Online version available in : http://arbitrer.fib.unand.ac.id

JURNAL ARBITRER

| 2339-1162 (Print) | 2550-1011 (Online) |

Article

# The Effect of Paper Reading Versus Screen Reading on the Inferential Reading Performance among University ESL Learners

*Musharraf Aziz*[1], *Ahdi Hassan*[2], *Omar Aljamili*[3]

[1]*University Utara Malaysia, Malaysia*
[2]*Global Institute of Research, Education and Scholarship, Netherlands*
[3]*University of Science and Technology, Yemen*

Submission Track

Received: February 21, 2024
Final Revision: March 25, 2024
Accepted: March 27, 2024
Available Online: April 5, 2024

Keywords

ESL learners, inferential comprehension, paper reading, reading mode, screen reading

Correspondence

E-mail: musharrafazizkaifi@gmail.com

A B S T R A C T

Contemporary ESL education focuses on screen-based reading, specifically in the context of university level learners. Also, a sudden shift from paper bound reading activities to screen-based tasks was inevitable in response to COVID-19 outbreak. In this regard, the case of inferential and advanced level reading among the young ESL learners of developing countries appears researchable because of the general low digital literacy of these learners. This study, therefore, aimed to investigate the effect of reading mode shift from paper to screen on the inferential comprehension performance among Pakistani ESL learners at university level. A total of 426 undergraduate learners were sampled from Bachelor of Science program in a well-known university in the city of Lahore, Pakistan. Inferential reading was conceptualized through Bloom's higher order thinking skills, that is, Analyzing, Evaluating and Creating skills therefore the reading comprehension test was based on these skills. Adopting sequential test administration, first the paper-based reading test, and later the screen-based reading was given. The obtained data were analyzed using Rendell's XCALIBRE and SPSS V. 26.0. The logit scale descriptives, learner ability (θ), and mean scores demonstrated that the learners performed significantly higher in the paper test as compared to screen test. Moreover, it was found that the effect of reading mode shift was larger in Creating skill. The findings may have significant implications for reading mode selection, enhancement of digital competence and effect of reading mode shift in developing countries with inadequate IT facilitation and digital literacy among ESL learners.

## I. INTRODUCTION

Technological advances have exponentially increased digital reading among university learners. Information technology has also permeated ESL education through a variety of digital reading devices while this trend has noticeably increased since 2007 (Biancarosa & Griffith, 2012). Specifically, young learners, in many ways, are required to execute screen-based reading activities. However, some significant critical questions remain unanswered while they are also investigable; How far can young learners cope with screen-based reading in developing countries in which most learners have relatively insufficient digital literacy? What impact has the sudden shift from paper to screen reading exerted on the inferential and critical reading of such learners? Keeping that in view, the current study investigated the effect of paper and screen reading on inferential comprehension of university level ESL learners in the post-COVID times. Though screen reading has not completely replaced paper reading, the increased use of screen reading devices such as laptops, digital tablets, smartphones and Kindle, is altering the paradigmatic perception

DOI: https://doi.org/10.25077/ar.11.1.85-95.2024

Under License of Creative Commons Attributioni-Non Commercial 4.0 International.

of reading comprehension among teachers and learners. Online reading environment, thus, offers a platform different from paper reading, providing ample room for investigating learners' cognitive performance in inferential reading that is processed through higher order thinking skills (Anderson et al., 2001). Mangen et al. (2013) contend that, superimposed by digital technology, university learners in developed educational contexts perceive reading as a screen phenomenon more than a paper activity. This perception and screen reading both have been boosted by the COVID-19 pandemic.

The COVID-19 restrictions reduced the activity of reading comprehension to its digitized version in the face of tactile precautions, lockdowns and possible future outbreaks (Bol, 2020; Wyse et al., 2020). Traditional reading underwent a massive mode shift from paper to screen. This unexpected transition posed certain reading challenges among university learners, particularly in the countries with low digital literacy. Most learners in developing contexts such as Pakistan, Iraq, Yemen and Syria, struggle with using advanced technology due to less exposure to instructional technology that is further linked to the lack of technological facilities in institutions. The challenges massively shifted from generation *X* to university learners posing issues with remote screen instructions, inferential reading comprehension, mind wandering and cognitive load. The digital transformation also raised questions of equivalence and disparity in assessing the learner performance (Pace et al. 2020). They argue that reading assessment mode has also changed with the suppressed use of paper reading assessment measures. Reading skill is important because generally university learners access the stock of knowledge through reading outside classroom. Further, inferential comprehension is required because it is the advanced form of comprehension that is vital for the university learner (Brookhart, 2015).

The remarkable shift from paper reading to screen reading and current hyper-phase of digitized reading comprehension invite investigation on the effect of this reading mode shift on the learner performance. Alongside, inadequate digital competence and poor facilitation may worsen the learning situation in developing education sectors. We, therefore, investigated the possible difference between Pakistani university learners' inferential comprehension performance based on digital interface and paper reading. In Pakistan, reading comprehension activities, prior to the COVID-19 outbreak, were largely paper-based from primary to tertiary education. Accordingly, the learners' e-learning habits were not adequately developed (Kanwal & Rehman, 2017). Moreover, limited literature exists on the possible performance alteration due to reading mode shift among Pakistani ESL learners. We, therefore, measured the inferential reading performance of Pakistani university learners who underwent massive reading mode shift.

Earlier studies have examined the effect of mode change on reading comprehension performance to investigate possible facilitation or impairment of comprehension process, differential technical features and effect of difference in textual layout. For instance, Kong et al. (2018) conducted a meta-analysis of 17 earlier investigations employing Robust Variance Estimation (RVE) to determine the difference in the learners' reading comprehension due to paper to screen transition. The findings demonstrated no significant difference due to the reading mode shift. They further found that the differences in the learner performance, observed before and after 2013, trended as a diminishing trajectory. These studies investigated reading comprehension as overall textual information processing without dividing it in lower order and higher order comprehension. For this reason, higher order thinking skills were not explicitly involved as standout variables in reading comprehension.

Regarding paper and screen reading mode, DeStefano and LeFevre (2007) contend that screen reading tends to increase the cognitive load because of embedded navigation models and hypertext features. This processual load influences the quality of comprehension sometimes to the extent of impairing it. Kalyuga (2012) found that effect of screen reading on the cognitive processing is attenuated when the syntactic features in the displayed text increase in complexity. Moreover, navigation during the process of comprehending a text and handling hypertexts may hinder the schematic arrangement of cognitive maps in the reader's mental dimension. This was supported by some recent studies. Ronconi et al. (2022) investigated the effect of change of text presentation mode involving reading comprehension, time, gender and performance calibration as factorial

variables. The linear mixed model was integrated with the possible interaction effect of gender. The findings indicated that the learner comprehension underwent certain cognitive changes at the level of core textual idea favouring paper reading.

Clinton (2019) conducted a meta-analysis of the studies ($n$ = 33) on paper to screen transitional effect on reading comprehension, that were conducted from 2008 to 2018. The consolidation was based on the variables of reading times, performance and performative calibration in metacognitive dimension. The random-effects model revealed that screen reading had a negative effect on the learner performance thereby favouring paper reading. Most studies in this area investigated adult reading while a significant case was investigated in Norwegian children. Norway digitized its standardized reading tests in 2016 that pre-require digital competence among the children of 10 to 14 years taking tests similar to PISA and PIRLS. Støle, Mangen and Schwippert (2020) conducted an experiment ($n$ = 1139) that measured the effect of paper and screen reading modes on the children's performance. The findings demonstrated that paper reading was better than screen activity. Another study conducted by Golan, Barzillai and Katzir (2018) on children aged 11 to 12 years, revealed that children preferred reading on screen however their comprehension performance was better while reading from paper.

The above inconsistent findings regarding the reading performance following reading mode change warrant further investigation. In this regard, several variables have been limitedly investigated such as reader's age, empirical mixed-method studies, eye-sight level of the learner/reader and sudden mode shift due to the COVID-19 outbreak.

The current study investigated the possible alteration in inferential comprehension performance based on paper and screen reading among Pakistani university undergraduates. A bi-modal assessment was designed adopting the higher order thinking skills (HOTS) hierarchy from revised Bloom's taxonomy (Anderson et al., 2001). They were namely *Analyzing*, *Evaluating* and *Creating* skills. Two inferential reading comprehension tests were designed and conducted in the first and last week of February 2023. Two hypotheses were developed on account of the reviewed literature on higher order reading assessment and mode shift. The first hypothesis was based on conceiving a mode shift effect in the learner performance, and required the comparative analysis of the scores of both reading modes. The second hypothesis was about the performance at HOTS levels, and required the comparative statistical analysis of the scores of *Analyzing*, *Evaluating* and *Reading* skills. *Creating* skill is the most advanced cognitive level in Bloom's taxonomy thus the reading researchers believe that it requires increased cognitive activity and comprehension than the other two skills (Anderson et al., 2001). This implied that the possible effect of reading mode shift may be inferentially larger in synthesizing new information from the given text than analyzing and evaluating it. These hypotheses are stated below:

**H1-*null***: Paper to screen reading mode shift in inferential comprehension does not demonstrate a mode shift effect in the mean score of the learners.

**H1-*alternative***: Paper to screen reading mode shift in inferential comprehension demonstrates a mode shift effect in the mean score of the learners.

**H2-*null***: The reading mode shift effect is not larger in *Creating* skill level as compared to *Analyzing* and *Evaluating* skills levels.

**H2-*alternative***: The reading mode shift effect is larger in *Creating* skill level as compared to *Analyzing* and *Evaluating* skills levels.

## II. METHODS

### Research design

This study employed quantitative research design drawing on the quantitative measures in data collection and analysis. A well-known university, 'The University of Lahore' (UOL) situated in Lahore City in Pakistan, was the research setting of this investigation. Lahore was chosen because it is considered the center of education and business in the country. Below are the investigation incumbents:

### Sample

A sample of 426 ESL learners was recruited from the two campuses of "The University of Lahore" in Lahore, Pakistan. We employed random probability sampling among the students of academic reading comprehension course (ARC-I-03) in Bachelor of Science in English Language and Communication (BSELC). This sampling strategy ensured that the learners shared a common academic background in being exposed to specific inferential reading comprehension

curriculum based on *Analyzing*, *Evaluating* and *Creating* skills. This assisted in mitigating wide sample heterogeneity in reading proficiency and the relevant external factors that could influence performance. The learners' ages ranged between 18 to 23 years according to the obtained demographic information. No ethnic presets were observed; however, these learners were Pakistani that incurred ethnic homogeneity. They had successfully passed their previous inferential reading examination therefore, score heterogeneity or large disparity was considered broadly absent.

**Instruments**

To assess via two modes of reading, two inferential reading tests were designed engaging higher order thinking skills; one text and test battery was for paper while the other was for screen assessment. Both tests were adapted from IELTS academic reading practice tests (IELTS, 2022) because IELTS tests are specifically designed for adult ESL learners. The learners took paper reading test one week earlier than the screen reading test. Both tests strictly followed the same map of higher order thinking skills with their interitem and overall test reliability determined two item analyses. The content and construct validity were constructing by following (i) the experts' comments, and relevant modifications until satisfactory validity was constructed, and (ii) simple-to-complex Bloom's HOTS hierarchy model in verbatim (Anderson et al., 2001).

Each test consisted of three texts with 1:1 ratio for the text and test. In selecting nature and length of the texts, the guidelines of Alderson et al. (2006) were used. He recommends using text of not more than 1000 words for inferential reading as crossing this threshold generally reduces the learner's cognitive processing of texts. Therefore, each text's length was < 600 words. The texts' readability was evaluated by the readability test proposed by McAlpine (2012) because this readability test was specifically designed for determining readability for adult ESL learners.

The paper text layout had linear and multimodal components while the screen text was in PDF form however, both followed the same layout criteria. In total, six different texts with different items were administered. *Analyzing* and *Evaluating* skills in inferential reading were tested by multiple-choice questions (MCQs) whereas *Creating* skill was evaluated by constructed response questions (CRQs). Creating skill items were different because many researchers (Brookhart, 2010; Kyllonen, 2017; Ramirez & Ganaden, 2008; Livingston, 2009) argue that creative skill should be assessed by CRQs because it requires the synthesis of detailed information that binary responses cannot accommodate. MCQs were used as Douglas, Wilson and Ennis (2012) highlight the efficacy of MCQs in examining complex cognitive processes related to analytical and evaluative reading. In both paper and screen reading tests, screen scrolling, and page turnovers were considered equivalent.

For administering a standardized reading comprehension battery, the HOTS hierarchy model was applied both at skills and subskills levels. *Analyzing*, *Evaluating* and *Creating* skills have the corresponding subskills that represent these three HOT skills. The MCQs and CRQs were distributed accordingly, with the number of CRQs less than that of MCQs due to the requirement of writing longer responses. The *Analyzing* and *Evaluating* subskills had 5 items each whereas *Creating* subskills had 3 items each, as shown in Table 1:

**Table 1.** Distribution of Items in the Inferential Reading Battery

| HOTS | MCQs | CRQs | Subskills |
|---|---|---|---|
| Analyzing | 15 | - | Differentiating, Organizing, Attributing |
| Evaluating | 10 | - | Checking, Critiquing |
| Creating | - | 9 | Generating, Planning, Producing |

The test was administered skills-wise to introduce the learners with the simple-to-complex hierarchical process as presented in Bloom's taxonomy. First, the text and test of *Analyzing* skill were given. They were followed by the texts and tests of *Evaluating* and *Creating* skills respectively. This purpose of adopting this testing scheme was two-fold; one, this was to administer texts and tests in the simple-to-complex hierarchical order. Two, it was to prevent memory decay in the learners because inferential reading comprehension tests generally are longer than lower order reading tests. The procedure of the research is described in Figure 1.

The learners were required to complete their responses in one hour and 30 minutes (i.e., 90 minutes) since this was the standard duration allotted to the learners for solving reading assessments in the participant university. The assessment yielded

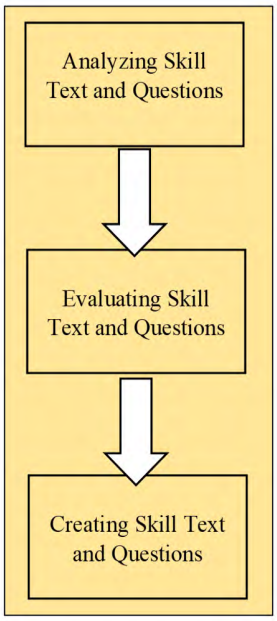



**Fig. 1 Skill-Wise procedure of reading text and test administration**

two core datasets of the solved test batteries in the form of paper reading test batteries (*n*= 426) and screen reading test batteries (*n*= 426). Each core datasets comprised 3 subsets of solved tests, based on *Analyzing*, *Evaluating* and *Creating* skills tests (see Table 1). Therefore, each subset of test batteries consisted of 426 solved tests, and each core dataset had 1,278 solved tests.

The researchers engaged the class instructors to mediate the test-taking procedure. This strategy was adopted to reduce the guest consciousness to a minimum among the respondents. The instructors were familiar with the test procedures of each reading mode because paper tests are traditionally used in Pakistan while screen reading tests were introduced during the COVID-19 outbreak. This prior familiarity of the learners with online testing since March, 2019 assisted in deciding not to diagnose the learners' digital competence for test-taking, however, only basic digital competence was required to take the screen-based test. The standard IBM computers, commonly referred to as desktop windows 10, were used for test administration. The tests were given in the UniLab and IT Compound in both campuses. The allocated response strategy for MCQ solutions was to use a mouse click whereas CRQs required the test-takers to use keyboards. Every solved test battery was programmed to generate an identity code, which was later assigned to each paper reading test before it was distributed accordingly. It was done so that both versions of each learner could later be matched during the data input. The paper reading tests were given on the standard A4 white papers. A checkmark (✓) was set to symbolize the selected answer.

Subsequent to collecting the data, the interrater reliability was determined by recruiting four raters, and assigning random two raters to evaluate each core dataset. Later, the scores of randomly evaluated 100 tests were matched to obtain a sufficient level of interrater reliability (α= 8.6) as recommended by Pallant (2020).

## III. RESULTS

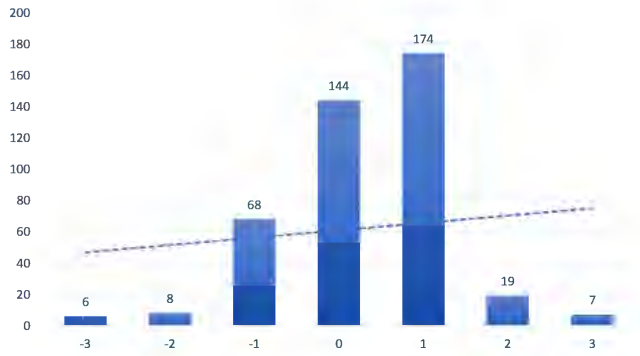
### Procedural Scaling of the Item Response Model

The reading performance was conceptualized through probabilistic model in accordance with the Item Response Theory (IRT). It assumed the model of continual equivalence between the latent parametric traits (i.e., *Analyzing*, *Evaluating* & *Creating* skills) and test items, to measure the respondents' ability (θ) against each trait. Therefore, the core objective of the evaluation was to locate the test-takers' position on the continuum of inferential reading trait. This theoretical stance provided the assumption of Local Independence for each item. For scaling, XCALIBRE (http://www.assess.com) and Rendell's IRT-Lab were used. XCALIBRE was used to compute MCQs by assuming dichotomous correct-incorrect pattern for the correct answers and distractors. IRT-Lab was incorporated for confirming the satisfactory scaling procedure especially in the graded response evaluation of Creating skill test items, which were CRQs. Each test-taker had two scores because of paper reading test and screen reading test. For comparing the reading scores of both modes, the linear scale was set on mean of 50 (*SD*= 10) for median probability.

An item-fit analysis was conducted to determine the overall model fit in XCALIBRE. For that matter, the items that were unfit due to unsatisfactory discrimination were extracted from the analysis. The discarding decision was made on the basis of the values of Chi-square and z-residuals, in relation with the *p* value, if it was equal or below 0.05, and if the ability (θ) principally fluctuated beyond -3 or +3, which are considered the normal range (Reise, 1990). However, items extraction was not a threat to the data because of the availability of ample number of items in the tests and normal data distribution.

Subsequent to scaling and data cleansing,

the frequency statistics (i.e., mean and standard deviation) demonstrated the learners' higher mean scores in paper reading test as compared to those in screen reading test. Alongside, the difficulty level of screen reading test was shown as higher than that of paper reading test that implied the emergence of an effect on the learner performance due to the reading mode shift. It favoured paper reading test mode as compared to the screen test mode. The logit scale showed profound implications of the computed probabilities favouring the learner performance in inferential comprehension on paper reading test. Figure 2 illustrates the standard deviation demonstrating the difference between paper and screen reading test.




***Note.*** Screen Reading Test= -ve; Paper Reading Test= +ve

**Fig. 2 Standard deviation of the learner performance in paper and screen modes**

Figure 3 shows the range from -3 to +3 representing the ability of the learners. On the scale, the negative values represent screen reading test while positive values refer to paper reading test. Further, positive values show higher performance whereas negative values symbolize screen lower performance. The scale shows that the performance of 144 learners, at zero scale, did not significantly deviate in paper and screen reading tests ($SD$ = 0.500 - 0.498). Meanwhile, 200 learners performed better in paper reading test as compared to 82 learners in screen reading test. Figure 2 shows a larger variation ($SD > 2$) in the number of the learner performance in paper reading test and screen reading test, favouring paper reading mode.

Figure 3 illustrates the number of the learners performing in both modes. As shown below, 73.75% learners performed higher in paper reading test whereas 26.25% learners performed higher in screen reading test. This indicated the learner performance percentage was higher in paper reading test than screen reading test.

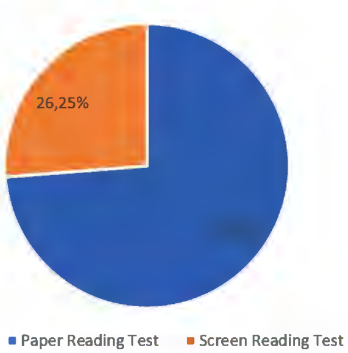



**Fig. 3 The learner performance percentage in paper and screen reading test**

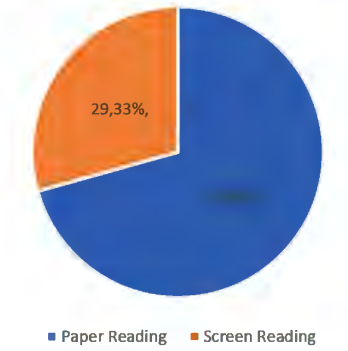



**Fig. 4 The percentage of the learners with fluctuated performance due to mode shift**

Figure 4 shows that in terms of higher performance, paper reading test had 70.66% learners whereas screen reading test had 29.33% learners. This indicated that the learner performance was higher in paper reading mode. Moreover, this supported H1-*alternative* while H1-*null* was rejected.

Subsequently, the learner percentage was determined based on their performance in higher order thinking skills individually. Figure 5 shows that the learners performed better in paper reading test (i.e., blue columns) than screen reading test (i.e., yellow). It also shows that the performance difference is larger in Creating skill than other skills. This confirms H2-*alternative* while H2-*null* is rejected.

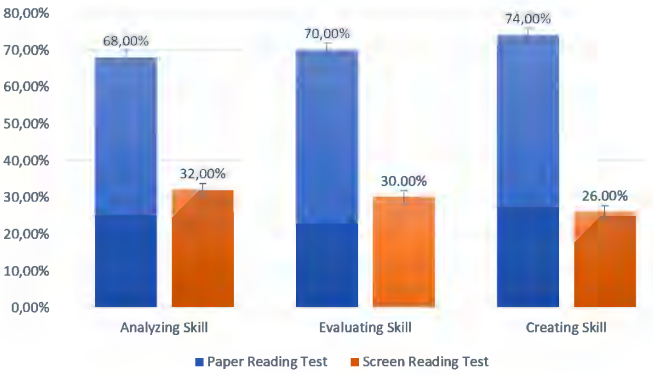



**Fig. 5 The learners performance in analyzing, evaluating and creating skills in paper and screen reading tests**

In inferential statistical analysis, a paired-sample *t*-test was conducted to determine the difference in the parallel score gain in overall inferential reading using the paper and screen modes. Table 2 presents the results:

**Table 2. T-test mean scores and standard deviation in inferential reading comprehension**

| Mode | Mean ($\bar{x}$) | Std. Deviation | *t* | Df | Sig. (2-tailed) |
|---|---|---|---|---|---|
| Paper | 58.76 | 17.14 | 2.49 | 424 | 0.012* |
| Screen | 52.69 | 16.90 | | | |

*$p < .05$

Inferential reading had three indicators namely *Analyzing*, *Evaluating* and *Creating* skills therefore, the possible performance difference in each indicator was determined conducting three paired-sample *t*-tests. Table 3 presents the results of the *t*-test conducted to determine the learner performance difference in *Analyzing* skill using paper and screen modes:

**Table 3. T-test mean scores and standard deviation in analyzing skill**

| Mode | Mean ($\bar{x}$) | Std. Deviation | *t* | Df | Sig. (2-tailed) |
|---|---|---|---|---|---|
| Paper | 10.67 | 2.73 | 2.14 | 424 | 0.035* |
| Screen | 8.86 | 2.39 | | | |

*$p < .05$

Table 3 demonstrates that, in *Analyzing* skill, the learners obtained a higher mean score in paper reading mode (x = 20.19, *SD* = 2.00) than screen reading mode (x = 17.25, *SD* = 2.21). This implies that there was a significant difference between *Analyzing* skill of the learners while inferentially comprehending the text through paper and screen modes, $t$ (198) = 3.38, $p = 0.02$ ($p < .05$).

**Table 4. T-test mean scores and standard deviation in evaluating skill**

| Mode | Mean ($\bar{x}$) | Std. Deviation | *t* | Df | Sig. (2-tailed) |
|---|---|---|---|---|---|
| Paper | 8.90 | 2.00 | 2.37 | 424 | 0.030* |
| Screen | 6.55 | 2.23 | | | |

*$p < .05$

In *Evaluating* skill also, the learner performance demonstrated a significant difference in the mean score of the paper mode (x = 20.19, *SD* = 2.00) and screen mode (x= 17.25, *SD* = 2.21), favouring the paper reading mode, $t$ (198) = 3.38, $p = 0.02$ ($p < .05$).

Table 5 shows that in *Creating* skill, the learners obtained a higher mean score in paper reading mode (x = 20.19, *SD* = 2.00) than screen reading mode (x= 17.25, *SD* = 2.21). This indicates

**Table 5. T-test mean score and standard deviation in creating skill**

| Mode | Mean ($\bar{x}$) | Std. Deviation | *t* | Df | Sig. (2-tailed) |
|---|---|---|---|---|---|
| Paper | 20.00 | 2.03 | 3.38 | 424 | 0.02* |
| Screen | 15.24 | 2.28 | | | |

*$p < .05$

a significant difference between the learner performance in *Creating* skill while reading via paper and screen modes, $t$ (198) = 3.38, $p = 0.02$ ($p < .05$).

Comparing the *t*-test results shows that the significance value ($p$) of the difference in the learner performance via paper and screen reading modes in *Creating* skill is greater than those of *Analyzing* and *Evaluating* skills. This provides an additional support to the descriptive findings on Hypothesis II.

## IV. DISCUSSION

The investigation of the effect of the reading mode shift on reading comprehension performance of the learners yielded significant findings. Specifically, the examination of the learners' performance in the paper and screen reading tests indicated that the learners performed better in paper reading test than screen reading test. This implies that ESL that most ESL learners in Pakistan, even at the post-COVID stage, tend to comprehend texts better via paper mode than screen mode. Several factors may be functioning behind this finding, for instance, it could be the learners' lack of digital literacy and experience with using educational technology. The findings align with Mangen et al. (2013) who found that print-reading cohort significantly outperformed screen-reading cohort. Further, in digital learning and cognitive research, a few researchers (Chen et al., 2014; Santos et al., 2019, Dolenc et al., 2015) found that screen scrolling has the potential of hindering the learner's reading flow. It implies that caution is required in mobilizing digital reading, in different contexts, because the success rate of the mode shift may depend on contextual traits, educational culture and several other factors such as digital literacy level. It also warrants further research on whether digital presentation of texts assists or impedes cognitive mapping among reader/learner (Mangen et al., 2013).

Brown (2023) also argued that text scrolling increases the cognitive load on the reader's mind. However, the effects may also relate with the

circumstances in which the readers are required to choose between the text presentation mode. For instance, in the COVID-19 times, screen reading was mostly the only option available to teachers and learners in order to avoid possible contamination due to reading involving tactile sources. Such restrictive circumstances pose related challenges such as the reduced learner performance, especially in the contexts with limited digital instructional technology and competence. Such socio-academic conditions need to be cautiously considered. In fact, the findings under discussion may plausibly be explained in terms of limited exposure of Pakistani ESL learners to digital reading.

Some previous investigations also found divergent outcomes indicating that the mode of reading does not significantly affect the precepts of inferential comprehension. For instance, Ocal et al. (2022) found that there was no significant difference in the learners' inferential comprehension while reading through paper and screen. In fact, Wolf (2022) argues that substantial benefits are linked with screen reading, such as multitasking, heightened information retention, attention locking and immersive cognitive engagement. He further asserts that screen reading is specifically beneficial in skimming activities.

The current findings also resonate with Jones et al. (2005) found that print reading assists comprehension and recall in the readers engaged with printed materials as compared to those reading from screens. A potential explanation for this divergence is the cognitive overload and eye strain on screens. The absence of digital distractions and the eye strain, seems to allow learners to effectively focus on textual content, leading to better recall. Furthermore, the tactile and spatial aspects of paper reading, including page turning and text palpability, can plausibly contribute to a more organized cognitive mapping. However, the second part of the same study found that longer memory is associated with screen reading. The researchers stated that was no readily available explanation of this result, which warrants further investigation on paper versus screen reading effects on comprehension.

It is also crucial to note the studies such as Wilson and Williams (2018) that report the effect of gaze-assisted Autopager, which furnishes 'change blindness' on digital screen due to its fade effect technique that does not require scrolling for page turnovers. This underscores the notion that a range of factors related to scrolling overload and lack of technological familiarity may be overcome by introducing automatic pagers thereby reducing the cognitive overload factor to a minimum. However, it seems that in developing contexts with limited educational facilities, the implications of reading mode change could be particularly significant, favouring paper reading. Many learners in these regions have limited access to digital devices and are thus more accustomed to paper learning. Consequently, the inherent advantages of paper reading, such as reduced cognitive load and improved information processing, may be even more pronounced in such contexts. This can be a possible explanation of the findings of the current study in the developing context of Pakistan though the growing prevalence of digital reading requires the reader adaptability to reading modes.

The *t*-test findings regarding the indictors of inferential reading comprehension (i.e., *Analyzing*, *Evaluating* & *Creating* skills) supported the undertaken hypotheses. There was found a consistent difference in performance not only in the overall inferential comprehension via paper and screen modes but also in these three indicators. The findings also revealed that the performance difference had a greater significance value in *Creating* skill than Analyzing and Evaluating skills, demonstrating a larger effect of transitioning from paper to screen mode. Moreover, it favoured paper reading mode for these ESL learners. It may be explained in terms of the requirement of expending more effort in understanding and managing the cognitive load in creative reading and writing activities than those of analytical and evaluative comprehension. A mode shift, that was also incurred in an emergency situation without much practice during COVID-19 pandemic, seems to have affected the learner performance. The larger effect also shows the possible involvement of textual interpretation and recording the answers via writing in paper mode and typing in screen mode as compared to ticking and crossing in MCQs, as was the case in the analytical and evaluative reading tests. Shaw and Weir (2007) argue that test-taking mode, particularly in writing that is usually a part of reading tests, may affect the test-takers' performance based on their digital literacy level. Keyboard literacy in computer-based tests may behave as a challenge to the learners in a context with low digital technology exposure and use (Zhi

& Huang, 2021). This may have been one of the reasons for the larger effect seen in the form of higher p-value in *Creating* test as compared to that of *Analyzing* and *Evaluating* skills tests.

## V. CONCLUSION

The current findings highlight the advantages of paper reading within the constraints of a specific context in which the learners generally have low digital competence because of the developing conditions of education and social upgradation. Thus, it hints at the complex interplay of cognitive processes, reading mode shift, and contextual factors. The findings underscore the superiority of the paper reading to screen reading within a controlled setting. These insights contribute to the ongoing discourse surrounding the multidimensional relationship between reading mode and comprehension. The study also emphasizes the necessity for a nuanced examination of cognitive, psychological, and technological dimensions. Furthermore, the differential exposure of learners in developing countries to screens due to limited resources amplifies the potential impact of reading mode on comprehension outcomes. Further exploration into the intricate interplay of reading mode, content type, individual preferences, and cognitive processes will be invaluable in understanding how learners engage with texts in an era of increasing digitalization.

## REFERENCES


Alderson, J. C., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (2006). Analysing tests of reading and listening in relation to the Common European Framework of Reference: The experience of the Dutch CEFR construct project. *Language Assessment Quarterly: An International Journal, 3*(1), 3-30. https://doi.org/10.1207/s15434311laq0301_2

Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J. and Wittrock, M.C. (eds.) (2001). *A taxonomy for learning and teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. Addison Wesley Longman. https://doi.org/10.1080/00461520903433562

Biancarosa, G., & Griffiths, G. G. (2012). Technology tools to support reading in the digital age. *The Future of Children*, 139-160. https://doi.org/10.1353/foc.2012.0014

Bol, T. (2020). *Inequality in homeschooling during the Corona crisis in the Netherlands*. First results from the LISS Panel. https://doi.org/10.31235/osf.io/hf32q

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. ASCD.

Brookhart, S. M. (2015*). How to make decisions with different kinds of student assessment data*. ASCD.

Brown, M. (2023). *Exploring Digital and Print Text Experiences of Adolescent Readers* (Published Doctoral Dissertation, Boise State University). ProQuest. https://doi.org/10.18122/td.2063.boisestate

Chen, G., Cheng, W., Chang, T. W., Zheng, X., & Huang, R. (2014). A comparison of reading comprehension across paper, computer screens, and tablets: Does tablet familiarity matter?. *Journal of Computers in Education, 1*, 213-225. https://doi.org/10.1007/s40692-014-0012-z

Clinton, V. (2019). Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of Research in Reading*, *42*(2), 288-325. https://doi.org/10.1111/1467-9817.12269

DeStefano, D., & LeFevre, J. A. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior, 23*(3), 1616-1641. https://doi.org/10.1016/j.chb.2005.08.012_

Dolenc, K., Aberšek, B., & Aberšek, M. K. (2015). Online functional literacy, intelligent tutoring systems and science education. *Journal of Baltic Science Education, 14*(2), 162. https://doi.org/10.33225/jbse/15.14.162

Douglas, M., Wilson, J., & Ennis, S. (2012). Multiple-choice question tests: a convenient, flexible and effective learning tool? A case study. *Innovations in Education and Teaching International*, *49*(2), 111-121. https://doi.org/10.1080/14703297.2012.677596

Golan, D. D., Barzillai, M., & Katzir, T. (2018). The effect of presentation mode on children's reading preferences, performance, and self-evaluations. *Computers and Education*, *126*, 346-358. https://

doi.org/10.1016/j.compedu.2018.08.001

IELTS. (2022). *The world's most trusted English test*. https://ielts.org/

Jones, M. Y., Pentecost, R., & Requena, G. (2005). Memory for advertising and information content: Comparing the printed page to the computer screen. *Psychology and Marketing*, *22*(8), 623-648. https://doi.org/10.1002/mar.20077

Kalyuga, S. (2012). Instructional benefits of spoken words: A review of cognitive load factors. *Educational Research Review*, *7*(2), 145-159. https://doi.org/10.1016/j.edurev.2011.12.002

Kanwal, F., & Rehman, M. (2017). Factors affecting e-learning adoption in developing countries–empirical evidence from Pakistan's higher education sector. *IEEE* Access, 5, 10968-10978. https://doi.org/10.1109/access.2017.2714379

Kong, Y., Seo, Y. S., & Zhai, L. (2018). Comparison of reading performance on screen and on paper: A meta-analysis. *Computers and Education*, *123*, 138-149. https://doi.org/10.1016/j.compedu.2018.05.005

Kyllonen, P. C. (2017). *Rethinking how we define and measure 21st century skills*. Oxford

Livingston, S. A. (2009). *Constructed-Response test questions: Why We Use Them; How We Score Them*. R&D Connections. Number 11. Educational Testing Service.

Mangen, A., Walgermo, B. R., & 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

Ocal, T., Durgunoglu, A., & Twite, L. (2022). Reading from screen vs reading from paper: does it really matter?. *Journal of College Reading and Learning*, 52(2), 130-148. https://doi.org/10.1080/10790195.2022.2028593

Pace, C., Pettit, S. K., & Barker, K. S. (2020). Best practices in middle level quaranteaching: Strategies, tips and resources amidst COVID-19. *Becoming: Journal of the Georgia Association for Middle Level Education*, 31(1), 2-13. https://doi.org/10.20429/becoming.2020.310102

Pallant, J. (2020). *Survival manual. A step-by-step guide to data analysis using SPSS*, 4(4). https://doi.org/10.4324/9781003117452

Ramirez, R. P. B., & Ganaden, M. S. (2008). Creative activities and students' higher order thinking skills. *Education Quarterly*, 66 (1).

Reise, S. P. (1990). A comparison of item-and person-fit methods of assessing model-data fit in IRT. *Applied Psychological Measurement*, 14(2), 127-137. https://doi.org/10.1177/014662169001400202

Ronconi, A., Veronesi, V., Mason, L., Manzione, L., Florit, E., Anmarkrud, Ø., & Bråten, I. (2022). Effects of reading medium on the processing, comprehension, and calibration of adolescent readers. *Computers and Education*, 185, 104520. https://doi.org/10.1016/j.compedu.2022.104520

Santos, A. I., Ferreira, C. M., Sá, M. J., & Serpa, S. N. F. D. (2019). Reading on paper and scrolling text on a screen in academic learning. *Academic Journal of Interdisciplinary Studies*, 8(3), 135-143. https://doi.org/10.36941/ajis-2019-0012

Scholarship Online. https://doi.org/10.1093/acprof:oso/9780199373222.003.0005

Shaw, S. D., & Weir, C. J. (2007). Examining writing: Research and practice in assessing second language writing (Vol. 26). Cambridge University Press.

Støle, H., Mangen, A., & Schwippert, K. (2020). Assessing children's reading comprehension on paper and screen: A mode-effect study. *Computers and Education*, 151, 103861.

https://doi.org/10.1016/j.compedu.2020.103861

Wilson, A. D., & Williams, S. (2018, June). Autopager: Exploiting change blindness for gaze-assisted reading. In Proceedings of the 2018 *ACM Symposium on Eye Tracking Research & Applications* (pp. 1-5). https://doi.org/10.1145/3204493.3204556

Wolf, M. (2020). Screen-Based Online Learning Will Change Kids' Brains. Are We Ready for That?. *The Guardian*. https://www.theguardian.com/commentisfree/2020/aug/24/deep-literacy-technology-child-development-reading-skills (Dated 24 August, 2020).

Wyse, A. E., Stickney, E. M., Butz, D., Beckler, A., & Close, C. N. (2020). The potential impact of COVID-19 on student learning and how schools can respond. *Educational Measurement: Issues and Practice*, 39(3), 60-64.

https://doi.org/10.1111/emip.12357

Zhi, M., & Huang, B. (2021). Investigating the authenticity of computer-and paper-based ESL writing tests. *Assessing Writing*, 50, 100548. https://doi.org/10.1016/j.asw.2021.100548