Developing Inferential and Monitoring Strategies in Adult EFL Reading: A Study from North Macedonia

Sandra Serafimovska [1], Karmen Pižorn [1]

[1] *University of Ljubljana, Faculty of Education, Kardeljeva ploščad 16, 1000 Ljubljana, Slovenia*

Abstract **- The reading comprehension of English as a foreign language is often measured in outcomes that value accuracy, with a relatively under-investigated body of literature on the individual cognitive and metacognitive processes involved. This study examines the use and development of inference and monitoring comprehension strategies by adult Macedonian learners in an online one-to-one instructional setting. It was based on a mixed-methods design whereby an experimental group was directly taught the reading strategies while control group was treated to regular instruction. Quantitative data was collected through pre- and post-reading comprehension tests, and qualitative data were obtained via think-aloud protocols. Results showed that while both groups demonstrated high baseline performance in literal comprehension, learners showed limited and largely implicit use of inference and monitoring strategies at baseline. After receiving the intervention, the experimental condition increased strategic awareness, systematic inferences, and demonstration of monitoring behaviours, such as self-questioning, rereading, and detection of comprehension breakdowns. These findings highlight the importance of process-oriented assessment and explicitly teaching specific strategies in adult English as a Foreign Language (EFL) reading and provide a pedagogical resource for strategy-oriented reading instruction.**

DOI: 10.18421/TEM153-81
https://doi.org/10.18421/TEM153-81

Corresponding author: Sandra Serafimovska,
University of Ljubljana, Faculty of Education, Kardeljeva ploščad 16, 1000 Ljubljana, Slovenia
Email: s.serafimovska990@gmail.com

Received: 04 February 2026.
Revised: 11 April 2026.
Accepted: 29 April 2026.
Published: 27 August 2026.

Keywords **- EFL reading comprehension, metacognitive reading strategies, inferential processing, adult EFL learners, North Macedonia.**

1. Introduction

Reading comprehension is a rich multilayered process. It is a process between decoding of written symbols, making meaning, and responding to those read symbols. Early theoretical models described reading as either a bottom-up process, governed largely by the recognition of words and syntactic thought, or a top-down process activated by readers' previous knowledge, anticipation, and predictions [1], [2]. Although these perspectives helped shape current interpretations of reading, they were criticized later for oversimplifying comprehension by focusing on a single processing pathway. Modern models view reading as an interactive process that requires lower-level linguistic skills and higher-order cognitive and metacognitive processes that are dynamically performed simultaneously [3]. The Simple View of Reading [4] viewed it as the product of decoding and linguistic comprehension, recognizing that comprehension involved both decoding and linguistic comprehension. Criticism of the model are that it provides a fairly static representation of reading, and that it understates readers' strategic and metacognitive engagement during comprehension [5]. Contemporary frameworks see reading as an active process of construction. The Construction–Integration model understands comprehension as an iteractive approach, as readers make inferences, apply textual information to understanding background knowledge and revise mental representations of the reading [6]. Likewise, the Active View of Reading assumes that readers operate on purpose as dynamic agents who use cognitive and metacognitive resources strategically to respond to the demands of written language, text complexity and comprehension challenges. Scarborough's Reading Rope extends that perspective by emphasizing that skilled reading begins with coordinated action of automatized skills with flexible higher order processes like inference generation and metacognitive regulation [7].

Across these models, a common assumption emerges: Successful reading comprehension is influenced by not just the language but also what readers infer as meaning and how well they monitor their understanding as they read.

1.1. Reading Skills and Reading Strategies in EFL Contexts

A core focus of research on reading is the distinction of reading skills and reading strategies. Reading skills are automatic procedures such as word recognition and syntactic parsing that build up slowly from repetition and do not involve much conscious effort [8], [9]. These skills are required for fluent reading, but not for deep comprehension. Reading strategies differ from the former two; they are intentional, goal-directed, and metacognitive actions taken by readers to meaningfully construct knowledge, mitigate comprehension issues, and manage understanding [10]. Inference generation, checking comprehension, rereading, and self-questioning require conscious control [11] and they are flexible and responsive, adaption is necessary to both task demands and text difficulty. Learners in EFL environments tend to acquire partial linguistic abilities, but not always maintain strategic control over the reading process. They might decode the texts and perform accurately on tasks addressing literal comprehension, but they have difficulty blending information across sentences, inferring or identifying comprehension breakdowns [12], [13], [14]. Explicit strategy teaching may support students' realisation of read process, thus making it easier to transfer strategies across texts and tasks [15]. However, the existing literature about EFL reading mainly focuses on outcome-based measures such as comprehension scores or self-report questionnaires that are unable to offer much explanation as to how strategies are implemented in reading. In response to this gap process-based approaches, such as think-aloud protocols, have been used to record readers' cognitive and metacognitive behaviour as it happens [16].

A core process related to reading comprehension is inference generation and monitoring comprehension. Inference allows readers to synthesize textual information with prior knowledge to build local and global coherence, and monitoring refers to learners' awareness of their comprehension and ability to manage reading behavior when reading comprehension fails. Failures to generate inference have been demonstrated to lead to fragmented understanding even in the presence of unbroken decoding and basic semantic processing [17]. On the other hand, monitoring comprehension works at the metacognitive level and can incorporate detecting inconsistencies, rereading, pausing to reflect and changing the methods.

Monitoring, in the Active View of Reading, is the key technique through which readers manage cognitive and metacognitive power in order to sustain comprehension [18]. In EFL, inference and monitoring are typically underdeveloped and are more dependent on performance on tests rather than reading strategically. Learners perform well on the literal comprehension tests without knowledge of inferential differences or comprehension shortcomings, indicating a contradiction between superficial comprehension performance and strategic reading ability.

1.2. The EFL Context in Macedonian EFL and the Motivation for the Current Research

In North Macedonia, English instruction and analysis with particular emphasis on reading assessment is heavily influenced by a standardized evaluation of English language, especially the national Matura examination, i.e., the final large-scale assessment of reading comprehension for the majority of learners. The English language reading test in the Matura exam mainly measures reading based on tasks that focus on the recognition of explicitly stated information, familiarity with vocabulary, as well as correct response selection. Thus, comprehension of reading is often conceptualised as something tangible rather than as constant cognitive and metacognitive. Internationally, North Macedonia's participation in the Programme for International Student Assessment (PISA), also points to the existence of continued difficulties of reading literacy. Even though PISA defines reading literacy as the ability to understand, use, assess and reflect on written books and texts for attainment of goals and social participation [19], recent cycles show that students face significant challenging tasking on tasks used to generate inferences, integrate information across texts, or evaluate meanings [20]. These results indicate an enduring divide between shallow and strategic comprehension of a text. In combination, Matura and PISA data suggest a reading profile where readers can display satisfactory performance on literal comprehension tasks and may show difficulty on higher-order reading tasks. But, while these evaluations do yield some important evidence on how students make sense of reading, how these students perceive an understanding breakdown, or how they control comprehension when the material becomes complicated, they provide little information about how readers make meaning when reading.

After secondary school, most of the systematic reading comprehension assessment and support disappear, which is also the case in North Macedonia, especially for independent learners who learn English for study in a private or non-traditional learning environment.

Reading is then taught and assessed in a test-based manner by which success is measured by how well you know the text and correct answers; adult EFL learners bring these reading practices into such contexts. As a result, strategic operations like inference generation and monitoring of comprehension are rarely discussed and are rarely or never fully explored. This instructional and assessment gap was the source of the current study. Observations of adult EFL learners in North Macedonia confirm that they decode texts effectively and perform excellently on explicit comprehension, but that they face difficulties to support deep reading and integrate information by inferring meaning or recognizing comprehension shortcomings. This mismatch between literal comprehension performance and strategic reading competence provided the primary motivation for the study. By focusing on inference and monitoring comprehension and by adopting a process-oriented research design, the study seeks to illuminate how adult EFL learners move beyond correct answers toward more regulated and meaningful reading.

2. Methodology

This study used a mixed methods experimental design, a way of collecting quantitative and qualitative data where it sought to explain the development of inferential and monitoring comprehension strategies in adult EFL reading. These quantitative data were used to compare reading performance pre- and post-instruction; qualitative data was used to provide process-oriented evidence of the learners' strategic behavior while reading. Research design consisted of experimental and control groups receiving explicit strategy instruction and regular reading instruction respectively. Based on these aims, the study addressed the following research questions:

RQ1. To what extent do adult EFL learners in North Macedonia demonstrate the use of inference and monitoring strategies during reading prior to explicit strategy instruction?

RQ2. Are there significant differences in inference and monitoring performance between learners who receive explicit strategy instruction and those who receive regular reading instruction?

2.1. Participants

This study used 50 adult EFL learners from North Macedonia, who for the duration of the study were enrolled in private, online one-to-one English lessons. Participants were randomly allocated through an experimental (n = 25) or control (n = 25) group to the intervention and controlled groups.

They were comparable in terms of age, gender, education level, education mix, work history, and learning of the English language. The participants ranged from between the ages of 24 and 53 years. The mean age was 34.6 years (SD = 7.74) in the experimental group, and 34.7 years (SD = 7.22) in the control group. Females as a whole represented 78% of the sample, which is consistent with typical enrollment data for those in private adult EFL teaching. All participants had a full-time job (100%). Regarding the participants' educational status, 68% of the total sample had master's degree, although those who had bachelor's degree were the few who participated in bachelor's study. Distributions of education were the same for both groups, demonstrating a very educated and well-balanced sample. Sample members hailed from diverse professional sectors such as business, information technology, administration, finance, education and other service-based sectors, and no sector dominated either one category. All participants attended English lessons in school for 10+ years with most of the years from official education (mainly by the national level). English proficiency self-report according to CEFR scale on levels of intermediate to upper intermediate (B1–B2) in reading, writing, listening and speaking. Each group reported different English use for their participants. Daily use was more common in the experimental group (36%) than in the control group (8%), whereas less frequent use was more frequent (72%) than in the experimental group (36%). There was little experience of standardized English proficiency tests for the students prior to the intervention. Most participants had no experience with international tests of standardized English and only a minority of them had previously been tested. This profile has implications in that participants' reading experience was largely molded by organized schooling and private instruction, rather than by standardized testing experiences. The study was free, voluntary and informed consent was obtained by all participants before the data collection.

2.2. Instructional Context and Procedure

Instruction took place in a one-to-one online format over a one-year period, with sessions conducted three times per week. This individualized setting allowed for close observation of learners' reading behavior and ensured consistency across instructional conditions. Both groups worked with comparable reading texts and tasks. At the beginning of the study (Phase 1), all participants completed a background questionnaire and a reading comprehension pre-test to establish baseline performance. The pre-test included items targeting both literal and inferential comprehension.

During the instructional phase, the experimental group received explicit instruction in reading strategies, with particular emphasis on inference generation and monitoring comprehension. Instruction involved strategy modeling, guided practice, verbalization of reasoning, and reflective discussion aimed at increasing learners' awareness of comprehension processes and regulation strategies. The control group followed regular reading instruction typical of private EFL settings, focusing on comprehension questions and vocabulary work, without explicit explanation or modeling of reading strategies. Instructional duration and frequency were equivalent across groups.

2.3. Instruments

Three main instruments were used for data collection. First, a background questionnaire was administered to collect demographic, educational, and language-learning information. Second, reading comprehension tests were administered before and after the instructional period to measure changes in reading performance, with particular attention to inferential comprehension. Third, think-aloud protocols (TAPs) were used to collect qualitative data on learners' cognitive and metacognitive activity during reading. During TAP sessions, participants were instructed to verbalize their thoughts while reading, including inferences, points of confusion, and monitoring behaviors.

2.4. Data Analysis

Statistical analysis of descriptive and inferential statistics in order to explore within-group and between-group differences was performed on quantitative data from the pre- and post-tests. The write-up of qualitative data on the think-aloud protocols were transcribed and then thematically analyzed, emphasizing evidence of inference generation, monitoring comprehension, and strategic regulation. This integrated nature of quantitative and qualitative findings enabled triangulation and a more complete account of how strategies were developed during the course of instruction.

3. Results

This section presents the results of the study, integrating quantitative data from reading comprehension tests given before and after the instructional period with qualitative data from think-aloud protocols (TAPs). The quantitative outcome is organized along strategy domain, reflecting the structure of the assessment instrument and the focus of the instructional intervention.

For each domain, scores are summed, enabling comparison between baseline and post-instruction performance within and between experimental and control groups. Alongside these outcome-oriented observations, TAP data are utilized to demonstrate how reading strategies were implemented while reading, and therefore, offer a process-oriented perspective on learners' inferential reasoning and monitoring behaviors. Collectively, the quantitative and qualitative results provide a more comprehensive understanding of reading comprehension, capturing both performance gains and strategic reading behavior.

3.1. Composite Scores of Inference Questions

Descriptive and inferential analyses were performed on participant information and inferencing skills before and after the educational intervention. The composite scores of the use of Inferencing strategy were separately analyzed in both experimental and control groups to assess intervention effects. Chi-square (χ^2) tests, Gamma, Kendall's Tau-b and post hoc residual analyses to explore significant associations and item-level contributions. Chi-square analysis revealed a statistically significant correlation between pre- and post-test scores for the control group ($\chi^2 = 17.5$, df = 9, p = 0.041), as well as for the total sample ($\chi^2 = 21.6$, df = 9, p = 0.010) as shown in Table 1.

Table 1. Results of χ^2 tests - Inference questions

Group		Value	df	p
experimental	χ^2	NaN	9	NaN
	N	25		
control	χ^2	17.5	9	0.041
	N	25		
Total	χ^2	21.6	9	0.010
	N	50		

Due to low expected frequencies, χ^2 for certain cells in the experimental group could not be calculated accurately; nevertheless, Gamma and Kendall's Tau-b analyses showed substantial positive relationships as shown in Table 2 and Table 3.

Table 2. Gamma coefficient values- Inference questions

			95% Confidence Intervals	
Group	**Gamma**	**Standard Error**	**Lower**	**Upper**
experimental	0.528	0.2369	0.0641	0.993
control	0.934	0.0694	0.7977	1.000
Total	0.765	0.1116	0.5466	0.984

Table 3. Kendall's Tau-b coefficient values- Inference questions

Group	Kendall's Tau-B	t	p
experimental	0.343	1.83	0.067
control	0.694	3.85	<.001
Total	0.533	4.16	<.001

More specifically, Gamma for the experimental group was 0.528 (SE = 0.237, 95% confidence interval: 0.064–0.993), which indicates a moderate positive correlation between pre- and post-test scores, whereas Kendall's Tau-b was close to being statistically significant ($\tau b = 0.343$, $t = 1.83$, $p = 0.067$). This was in sharp contrast to the control group (Gamma = 0.934, SE = 0.069, 95% CI: 0.798–1.000; $\tau b = 0.694$, $t = 3.85$, $p < 0.001$) and to the overall sample (Gamma = 0.765, SE = 0.112, 95% CI: 0.547–0.984; $\tau b = 0.533$, $t = 4.16$, $p < 0.001$). Post hoc residual analysis found that participants who scored 4 post-intervention had the largest contributions to overall difference of outcome, which indicates moderate improvement in information/inferencing skills after the educational intervention. In the control group, residuals differed more significantly at scores lower than 4 and higher than 5 and this was seen with greater variability within the control condition. For instance, the participants in the experimental group showed marginal improvements between pre- and post-test measurements as evidenced in the descriptive statistics, with some movement to the intermediate grades. The control group had improvement in a wider spread, although there was no discernable order at higher scores. In summary, the analyses indicate statistically significant difference ($p<0.05$) between pre- and post- tests. These findings indicate that the educational intervention resulted in effects in information and inferencing skills in the experimental group, albeit both groups showed some pre-post change. The placement of Gamma and Kendall's Tau-b show a strong association between pre- and post-test scores, with the experimental group having a moderate positive correlation and the control group a more significant correlation.

3.2. Composite scores of Monitoring Comprehension

Descriptive and inferential analyses were performed to assess participants' use of Monitoring comprehension strategy prior to and following the educational intervention. The composite scores for Monitoring comprehension were looked at separately for the experimental and control groups to see if the intervention had any effects. Then, chi-square (χ^2) tests, Gamma, Kendall's Tau-b, and post hoc residual analyses were used to find important connections and item-level contributions. Chi-square analysis revealed a statistically significant correlation between pre- and post-test scores for the experimental group ($\chi^2 = 14.9$, $df = 4$, $p = 0.005$) and for the total sample ($\chi^2 = 13.0$, $df = 4$, $p = 0.011$) as shown in Table 4.

Table 4. Results of χ^2 tests - monitoring comprehension questions

Group		Value	df	p
experimental	χ^2	14.9	4	0.005
	N	25		
control	χ^2	NaN	4	NaN
	N	25		
Total	χ^2	13.0	4	0.011
	N	50		

Chi-square could not be reliably computed for the control group due to low expected frequencies. However, Gamma indicated a strong positive association in the experimental group ($\gamma = 0.714$, SE = 0.215, 95% CI: 0.294–1.000), a weak positive association for the control group ($\gamma = 0.368$, SE = 0.362, 95% CI: –0.341–1.000), and a moderate positive association for the total sample ($\gamma = 0.548$, SE = 0.205, 95% CI: 0.146–0.950) as shown in Table 5.

Table 5. Gamma coefficient values - monitoring comprehension questions

			95% Confidence Intervals	
Group	**Gamma**	**Standard Error**	**Lower**	**Upper**
experimental	0.714	0.215	0.294	1.000
control	0.368	0.362	-0.341	1.000
Total	0.548	0.205	0.146	0.950

Kendall's Tau-b showed a significant positive association for the experimental group ($\tau b = 0.385$, $t = 2.021$, $p = 0.043$), a non-significant association for the control group ($\tau b = 0.188$, $t = 0.967$, $p = 0.333$), and a significant positive association for the total sample ($\tau b = 0.292$, $t = 2.177$, $p = 0.029$) as shown in Table 6.

Table 6. Kendall's Tau-b coefficient values - monitoring comprehension questions

Group	**Kendall's Tau-B**	**t**	**p**
experimental	0.385	2.021	0.043
control	0.188	0.967	0.333
Total	0.292	2.177	0.029

Post hoc residual analysis for the experimental group revealed that participants scoring 2 pre-intervention contributed to increases in post-test scores of 3 and 4, indicating a shift toward higher metacognitive skills after the intervention. Residuals in the control group were smaller and distributed, suggesting minimal change across pre- and post-test scores.

Descriptively, participants in the experimental group demonstrated observable improvements from pre- to post-test, with a larger proportion achieving higher monitoring and metacognitive scores, whereas the control group showed limited gains. In summary, the analyses indicate statistically significant difference ($p<0.05$) between pre- and post- tests. These findings suggest that the educational intervention had a positive effect on metacognitive and monitoring skills, with Gamma and Kendall's Tau-b supporting the strength of the observed associations.

3.3. TAPS Results

The study used Think-Aloud Protocol (TAP) data to conduct a real-time cognitive and metacognitive activity of learners while reading, focusing heavily on inference generation and the tracking of comprehension skills. Pre-test and post-test TAP data were analysed thematically and compared between the experimental and control groups.

3.3.1. Inference in Think-Aloud Data

In the pre-test TAPs, inference generation was poor and varied between both groups. Learners focused on decoding single words or paraphrasing sentences without integrating the input with the whole text. In summary, when conclusions were drawn, they were generally tentative in nature and were limited to specific sentences and no coherent interpretations were made. "This word has a meaning about money but may have to do with business". (Pre-test, Experimental group). "I know this, but I don't know why it's here". (Pre-test, Control group). Learners explicitly refrained from inferential reasoning and tended to use surface-level understanding in many cases. "It doesn't really say it, so I'm not sure. I'll just jump to the next sentence". (Pre-test, Experimental group). "Maybe we should not understand everything". (Pre-test, Control group).

Post-test TAPs in both groups showed an increase in their inferential activity. They reported participating in more attempts to infer implicit meaning, relating ideas from sentence to sentence, and accessing earlier parts of the text more or less to help explain their ideas. "I don't write it out, but I think this decision was given already from the last paragraph". (Post-test, Experimental group). "I don't feel they explain it here, but they're discussing a comparable case earlier — so I think it's connected". (Post-test, Control group). Moreover, some of the learners started to justify their inferences, suggesting greater insight into the inferential process as well. "They mentioned the deadline before, and now they talk about pressure". (Post-test, Experimental group). But the generation of inference was often localized. Learners seldom consulted or revised prior inferences between passages and inferential reasoning was not consistently maintained across longer passages. This trend was found in both groups and corresponds to the quantitative findings indicating high overall improvement, with no significant between-group discrimination.

3.3.2. Monitoring Comprehension in Think-Aloud Data

Behaviors that were monitored exhibited larger group differences. Participants in the two groups had relatively low levels of metacognitive awareness pre-test. Learners constantly read despite reading in a state of uncertainty, and hardly paused to reflect or regulate what they were reading. “I don’t understand this sentence well, but what might work next?” (Pre-test, Experimental group). “This part has lost everything, but I’m simply going to say so”. (Pre-test, Control group). In others, learners downplayed problem with comprehension instead of confronting them. “It’s hard, but I feel like it’s OK if I don’t get it all”. (Pre-test, Control group).

The monitoring behaviors were made explicit and systematic after the post-test, especially in the experimental group. Learners were increasingly verbalizing comprehension problems, pausing to assess the meaning of tasks, and beginning repair strategies like rereading or paraphrasing. “Sorry, this doesn’t seem right to me. I wanted to read it again because it matters for the entire paragraph.” (Post-test, Experimental group). “I believe I misunderstood this part. I will return to that, original sentence.” (Post-test, Experimental group).Students in the experimental group also showed anticipatory monitoring and anticipated possible difficulties before comprehension was broken. “This looks like an extremely difficult thought, so here again I should read with more care.” (Post-test, Experimental group).By contrast, although a small improvement was noted in the control group as well, monitoring behaviors tended to be less persistently documented and articulated. “I don’t know here but possibly it will be explained later.” (Post-test, Control group). “I think I get the main idea if some bits are not clear.” (Post-test, Control group).

3.3.3. Summary of TAP Patterns

Although inference generation across TAP data increased in both groups, this inference generation was frequently still fragmented and restricted to local segments of the text. However, the monitoring comprehension showed more distinctly qualitative differences between groups, in both cases with the experimental group demonstrating more metacognitive awareness, more frequent recognition of comprehension breakdowns and more intentional regulatory behavior during reading. These qualitative trends fit the quantitative observation and exemplify how patterns of reading behavior developed in the reading process.

4. Discussion

In this current study, inference and monitoring comprehension strategies for adult EFL reading were evaluated using a mixed-methods design employing pre- and post-test methods complemented by think-aloud protocols. This study sought to explore how adult learners constructed and regulated meaning in reading in a way that was not solely aligned with correct answers by merging results-based and process-based data.

Inference development: Quantitative data revealed that both the pre-test and post-test inferential performance were improved in the experimental group and their control counterparts, with a significant difference from 0–1 correct to 2–3 correct scores on inference items. These findings indicate that in the absence of explicit strategy-focused instruction, adult EFL learners are able to incrementally acquire inferential accuracy over long periods of time. However, TAP results add significant nuance to this result. While students would increasingly be looking for implicit meaning in the post-test, inferential reasoning was often localized, tentative, and poorly integrated across the text. Learners were much more dependent on text and sentence structures/cues that were isolated, rarely revisited/re-made inferences as they read on. This trend occurred in both groups and supported the lack of substantial between-groups differences in quantitative inference findings.

The results are also theoretically supported based on the Construction–Integration model which highlights inference as central to constructing textual coherence [6] . Although the inference items that most people succeeded in were relatively straightforward because participants used these measures of success on inference items, the TAP data indicated that many inferences that learners managed reached the local coherence level, with little internal integration into a global mental model of the text. This also echoes prior research that suggests the accuracy of inference alone does not guarantee a deep understanding, especially in EFL contexts where linguistic constraints restrict more tactical flexibility. Importantly, the inference items of test questions in this study were scored in the study as correct/incorrect, thus presumably having a one-sided (correct/incorrect) score of the types of inferential reasoning, this may have influenced its sensitivity to qualitative variations in inferential processing in qualitative data. Hence, the amalgamation of quantitative and TAP data points to a critical distinction between generating right and wrong inferential answers and sustained inferential reasoning while reading that is critical to a reading task.

Monitoring comprehension: Unlike inference, monitoring comprehension provided a clearer distinction between the groups even in the post-test period. The quantitative data revealed that an increased percentage of members of the experimental group attained higher cumulative monitoring scores, indicative of higher metacognitive awareness. These results were substantially corroborated by TAP data. Post-test TAPs from the experimental group exhibited repeated explicit monitoring behaviors such as verbalizing of uncertainty, rereading, self-questioning and selective control of reading pace. Students were aware of comprehension breakdowns and actively sought to address them. In comparison, while some improvement could be observed in the control group, monitoring behaviors tend to be less systematic and less explicitly stated.

These findings closely mirror the Active View of Reading, which posits reading as a purposeful, self-regulated endeavor, where the reader purposefully modulates cognitive and metacognitive resources [5]. Monitoring involves conscious awareness and control, so it responds to clear instruction and guided practice quite well. Whereas inference emerges implicitly during a process of exposure to content and task familiarity, monitoring seems to benefit most directly from instructional strategies that emphasize strategic reflection and regulation.

More information on the TAP data further suggests the facilitating role of monitoring in generating inference. Learners who consistently monitored their understanding were more prone to stop, re-consider the meaning, and make inferential connections, while learners who did not monitor frequently moved on seamlessly in the face of uncertainty. This is in line with previous research into monitoring as a major determinant of effective comprehension regulation in both L1 and EFL reading contexts.

Collectively, the result suggests that although inferential accuracy may grow over time, monitoring comprehension reflects a more sensitive indicator of long-term reading strategy gains in response to direct instruction. This mix of statistical and qualitative evidence serves to emphasize the need to conceptualize reading as a cognitive, as well as meta-cognitive activity, instead of simply as a process of obtaining (right) answers. For adult EFL learners in North Macedonia, aiding the adoption of monitoring strategies may particularly be essential for the development of deeper, more flexible reading understanding at the level of comprehension beyond performance in a test.

5. Conclusion

The present study aimed to investigate inference and comprehension monitoring strategies in adult EFL reading by integrating quantitative measures of performance with process-based insights derived from think-aloud protocols. The study examined how meaning construction and the regulation of meaning unfold during reading, and thus moved beyond accuracy-based outcomes to explore what reading comprehension looks like in an adult EFL context. The findings suggest that inferential accuracy may show improvement through repeated exposure to reading tasks, even in the absence of explicit strategy instruction. However, qualitative evidence indicates that such gains do not necessarily reflect sustained or integrated inferential reasoning. In contrast, comprehension monitoring emerged as a more sensitive and reliable indicator of strategic development, particularly when supported through explicit instruction. Learners who received targeted strategy instruction demonstrated increased awareness of comprehension breakdowns and greater intentionality in regulating their reading processes. These findings highlight a critical distinction between performance-oriented outcomes and strategic reading competence. In educational contexts such as North Macedonia, where reading assessment remains largely standardized and outcome-driven, learners may achieve acceptable test scores while lacking robust metacognitive control over comprehension. The present study therefore highlights the importance of prioritizing strategic awareness and regulation as central components of reading development, rather than focusing exclusively on correct responses. From a pedagogical perspective, the results point to the value of explicitly teaching comprehension monitoring strategies as a means of fostering deeper engagement with texts and promoting self-regulated reading. Instruction that supports learners in noticing, evaluating, and responding to comprehension difficulties appears to play a crucial role in moving beyond surface-level understanding toward more flexible and adaptive reading practices. Methodologically, the study demonstrates the strength of combining think-aloud protocols with traditional assessment measures. The integration of process-oriented data provided insights into learners' strategic behavior that would not have been visible through performance scores alone, reinforcing the need for research approaches that capture reading as it unfolds in real time.

At the same time, several limitations should be acknowledged. The study was conducted with a relatively small sample of adult EFL learners engaged in one-to-one online instruction in North Macedonia, which may limit the generalizability of the findings to broader instructional contexts. Additionally, the focus on inference and comprehension monitoring represents only a subset of the broader repertoire of reading strategies, leaving other higher-order processes such as evaluation, synthesis, and the strategic use of background knowledge unexplored. The individualized instructional setting may also have amplified the effects of explicit strategy instruction, particularly in relation to monitoring processes. These limitations point toward important directions for future research. Further studies could examine the development of reading strategies across larger and more diverse samples, as well as in group-based or classroom settings, to determine whether similar patterns of strategic development emerge. There is also a need to investigate how multiple strategies interact and evolve over time, particularly in digitally mediated reading environments. Longitudinal and multi-method research designs may offer valuable insights into how adult EFL learners transition from accuracy-driven reading approaches toward more flexible, self-regulated models of comprehension.

In sum, the present study contributes to a growing body of research that reconceptualizes reading as an active, strategic, and self-regulated process. Developing adult EFL readers who are capable of monitoring, regulating, and adapting their comprehension requires instructional and assessment practices that move beyond the evaluation of isolated correct answers and instead recognize the dynamic nature of meaning-making in reading.

References:

[1]. Goodman, K. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist, 6*, 126-135.

[2]. Gough, P. B. (1972). One second of reading. *Language by ear and by eye*, 331-358. MIT Press.

[3]. Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.

[4]. Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and special education*, *7*(1), 6-10.

[5]. Duke, N. K., Ward, A. E., & Pearson, P. D. (2021). The science of reading comprehension instruction. *The Reading Teacher*, *74*(6), 663-672. Doi: 10.1002/trtr.1993

[6]. Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

[7]. Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. *Handbook of early literacy research*, *1*, 97–110. Guilford Press.

[8]. Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific studies of reading*, *11*(4), 357-383.

[9]. Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.

[10]. Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The reading teacher*, *61*(5), 364-373.

[11]. Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Lawrence Erlbaum Associates.

[12]. Cain, K., & Oakhill, J. (2008). *Children's comprehension problems in oral and written language: A cognitive perspective*. Guilford Press.

[13]. Oakhill, J., Cain, K., & Elbro, C. (2014). *Understanding and teaching reading comprehension: A handbook*. Routledge.

[14]. Grabe, W., & Stoller, F. L. (2019). *Teaching and researching reading* (3rd ed.). Routledge.

[15]. Pressley, M. (2006). *Reading instruction that works: The case for balanced teaching* (3rd ed.). Guilford Press.

[16]. Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (rev. ed.). MIT Press.

[17]. Oakhill, J., Berenhaus, M. S., & Cain, K. (2015). Children's Reading Comprehension and Comprehension Difficulties. *The Oxford Handbook of Reading*. Oxford University Press.

[18]. Duke, N. K., Ward, A. E., & Pearson, P. D. (2021). The science of reading comprehension instruction. *The Reading Teacher*, *74*(6), 663-672. Doi: 10.1002/trtr.1993

[19]. OECD. (2019). *PISA 2018 results (Volume I): What students know and can do*. OECD Publishing.

[20]. OECD. (2022). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing.