Journal of Basic Education Research
Vol. 6, No. 3, September 2025, pp. 461~469
ISSN: 2716-3725, DOI: 10.37251/jber.v6i3.2102

Reading Literacy Assessment in Philippine Basic Education: A Cross-Sectional, School-Based Phil-IRI Study

Julie Ann V. Balisoro[1] , Aaron A. Funa[2] , Renz Alvin E. Gabay[3]
[1, 2, 3] Graduate School, Sorsogon State University, Sorsogon City, Sorsogon, Philippines
[1] Sorsogon National High School, Sorsogon City, Sorsogon, Philippines

Article Info

Article history:

Received Aug 8, 2025
Revised Sep 22, 2025
Accepted Sep 26, 2025
OnlineFirst Sep 28 , 2025

Keywords:

Expectancy-Value Theory
Reading Comprehension
Reading Proficiency
School-Based Assessment
Sex Difference

ABSTRACT

Purpose of the study: To examine whether age, sex, and grade level are associated with Philippine Informal Reading Inventory (Phil-IRI) performance in junior high school.

Methodology: Cross-sectional, ex post facto analysis of n = 200 Grade 7–10 students (SY 2024–2025). Phil-IRI (Filipino) post-test scores were compared using Mann–Whitney U (sex) and Kruskal–Wallis H (age bands; grade levels), $\alpha = .05$.

Main Findings: Sex was associated with Phil-IRI scores: females outperformed males (Mdn = 26 vs 19; $p < .001$). No differences were found by age ($p = .087$) or grade ($p = .513$). Descriptives showed a small, lower-performing oldest age band and non-monotonic medians across grades, indicating variation occurs mainly within rather than between grades.

Novelty/Originality of this study: Phil-IRI outcomes in this cohort relate to sex, not to age or grade. Schools should prioritize within-grade, profile-based supports using Phil-IRI decision levels and implement boy-responsive engagement, while monitoring older learners who remain below independence.

Corresponding Author:
Aaron A. Funa,
Graduate School, Sorsogon State University,
Sorsogon City, Sorsogon, Philippines.
Email: funa.aaron@sorsu.edu.ph

1. INTRODUCTION

Reading comprehension underpins content learning, vocabulary growth, and higher-order thinking from primary through secondary grades; recent syntheses link purposeful reading activities and sustained motivation with measurable gains in comprehension [1], [2]. Yet system indicators remain stark. In PISA 2018, the Philippines' mean reading score was 340, far below the OECD average of 487, and most Filipino 15-year-olds fell below the minimum proficiency benchmark; only 19% reached at least Level 2 [3]. In PISA 2022, girls outperformed boys in reading by 35 points in the Philippines, mirroring global patterns and signaling subgroup disparities that local diagnostics may capture [4]. Globally, about four in ten students still do not achieve minimum proficiency in reading by the end of primary, underscoring the urgency of targeted, evidence-based supports [5].

Within this context, the Philippine Informal Reading Inventory (Phil-IRI) operates as a classroom-based diagnostic across three modalities oral reading (word recognition/fluency), silent reading (independent comprehension), and listening comprehension classifying students at independent, instructional, or frustration levels to guide remediation and progress monitoring [6]. In practice, schools follow a four-stage cycle: (1) group screening to flag learners at risk; (2) individualized graded-passage pretests to pinpoint profiles; (3) targeted intervention (small-group or individualized); and (4) post-tests to verify movement across levels. Results are reported at learner, class, and school levels and feed into Every Child A Reader Program (ECARP) planning and School Improvement Plans (SIPs), so Phil-IRI data trigger tiered supports rather than static labels [7], [8]. Although designed for elementary grades, recent division and regional directives have extended Phil-IRI to junior high (Key Stage 3): for SY 2024–2025, divisions instructed schools to administer Phil-IRI (English and Filipino) in Grades 7–10 using the 2018 manual and newly developed junior-high materials, enabling consistent benchmarking from upper primary into lower secondary [9]-[11]. This architecture provides a practical lens to distinguish who benefits from universal whole-class routines versus who requires targeted or intensive intervention and to verify movement from frustration to instructional or independent levels over the school year.

Despite Phil-IRI's widespread use, evidence on junior-high cohorts remains limited, especially on whether age, sex, and grade are associated with post-test outcomes once Phil-IRI was extended to Grades 7–10. This study contributes school-level, decision-oriented evidence by analyzing an entire junior-high cohort in a large Philippine public school and interpreting subgroup patterns through established reading frameworks. The goal is to move beyond descriptive compliance reporting toward signals that can inform remediation, monitoring, and School Improvement Plans.

We anchor interpretation in the Simple View of Reading (SVR), which models comprehension as the joint product of word recognition and language comprehension; subgroup differences often reflect imbalances in these components [12], [13]. This mapping aligns with Phil-IRI modalities: oral reading indexes word recognition/fluency, listening comprehension approximates language comprehension, and silent reading integrates both. Accordingly, profiles such as "poor decoder" (weak oral reading with relatively stronger listening) and "poor comprehender" (adequate oral reading with weak listening) are theoretically expected, and demographic contrasts can be interpreted through this lens. Complementing SVR, Scarborough's Reading Rope highlights the intertwining of lower-level (phonology, decoding, fluency) and higher-level strands (vocabulary, background knowledge, verbal reasoning), predicting specific Phil-IRI weaknesses rather than uniform deficits when strands develop unevenly [14]. The Lexical Quality Hypothesis likewise holds that precise, well-integrated word representations facilitate comprehension, implying that exposure and instruction—which can differ by age, grade, and sex—will affect Phil-IRI levels [15].

Despite extensive international work on reading motivation and component skills, Philippine Phil-IRI studies have largely focused on elementary cohorts or single-grade snapshots, with limited analysis of how age, sex, and grade jointly relate to junior high outcomes [12]–[18]. As schools extend Phil-IRI use into Key Stage 3, evidence on these demographic associations is needed to inform screening and supports. Because effort and persistence mediate skill growth, we also adopt an expectancy–value perspective: learners engage more when they believe they can succeed (expectancy), value the task (intrinsic and utility value), and perceive manageable costs [16]. International syntheses show that girls report higher reading enjoyment and time on task, which partly explains observed achievement gaps; local assessments likewise document a female advantage [16], [17]. Under this account, sex differences in Phil-IRI may reflect both component-skill differences (decoding, language) and motivation-linked engagement and practice. Age and grade effects are expected to be nonlinear: as curricular demands rise, students with weaker decoding or language foundations may experience higher perceived cost and reduced expectancy, dampening gains despite additional years of schooling. These frameworks therefore yield testable expectations for Phil-IRI distributions (e.g., higher independent-level proportions among girls; non-monotonic changes across grades when cohorts include repeaters with persistent weaknesses) and justify empirical tests of demographic–outcome links rather than uniform assumptions [12]-[18].

To address this gap, we conduct a school-based cross-sectional analysis that simultaneously examines age, sex, and grade as correlates of Phil-IRI performance among Philippine JHS students. The study's novelty lies in linking a national screening tool to adolescent literacy profiling at scale, and its urgency follows from persistent proficiency shortfalls that demand targeted, data-informed supports [5]. We build on evidence that sex differences in reading often favor girls and that local Phil-IRI snapshots reveal uneven outcomes requiring targeted support [18], while recognizing persistent global proficiency shortfalls that heighten the urgency of data-informed interventions [5]. Accordingly, our objectives are to (a) describe the demographic profile of respondents; (b) compare Phil-IRI distributions across their demographic profile; and (c) determine significant difference between their demographic profile and their Phil-IRI scores..

2. RESEARCH METHOD

This study employed a cross-sectional, analytic, ex post facto design using routine school-based assessment data to test whether age, sex, and grade level are associated with reading performance as classified by the Phil-IRI. Cross-sectional analytic designs estimate associations among variables measured at a single time point without researcher manipulation—appropriate for administrative datasets and for questions about subgroup differences [19]-[21]. The design is ex post facto because the putative "exposures" (demographics) are naturally occurring and cannot be assigned; analysis therefore focuses on observed group differences rather than causal effects [22], [23]. Analyses used de-identified administrative records from one large public junior high school during the target school year; procedures and reporting followed good-practice guidance for observational cross-sectional studies [24].

We analyzed records from 200 Grade 7–10 students enrolled in a large Philippine public junior high school during SY 2024–2025 who completed the regular Phil-IRI administration. The cohort was gender-balanced (female = 101, 50.5%; male = 99, 49.5%), supporting unbiased subgroup comparisons. Grade-level representation was uniform (50 students per grade for Grades 7–10, each 25.0%), enabling like-for-like comparisons across junior high. Age was distributed as follows: 11–12 years (n = 37, 18.5%), 13–14 (n = 101, 50.5%), 15–16 (n = 58, 29.0%), and 17–18 (n = 4, 2.0%). Eligibility required enrollment in Grades 7–10 during the testing window and availability of official Phil-IRI (Filipino) results from school records; demographic data (age, sex, grade) were taken from the same records. The sample is near 1:1 sex ratio and equal grade strata provide balanced power for planned non-parametric comparisons by sex and grade, while the age-band distribution allows sensitivity checks for potential age–grade mismatch effects.

We used the Phil-IRI as the primary instrument. The Phil-IRI is a nationally mandated classroom assessment that profiles students' reading performance in English and Filipino [6]. For each modality, learners are classified at one of three decision levels independent, instructional, or frustration which are intended to guide instructional planning and remediation [6]. Functionally, Phil-IRI serves a role similar to informal reading inventory protocols used internationally, but with standardized procedures and reporting embedded in the Philippine public school system. Results are aggregated for learners, classes, and schools and feed into the ECARP and school improvement planning, ensuring that data inform tiered supports rather than remain descriptive labels [8], [9]. In this study, we analyze the official Phil-IRI decision levels as ordinal outcomes to compare distributions by age, sex, and grade.

We secured approvals from the school head and division office, complied with institutional ethics requirements. We did not introduce any instructional treatment. All data came from the school's routine administration of the Phil-IRI during the division's scheduled window for the target school year. We ensured that Phil-IRI was administered by trained teachers in accordance with the 2018 Phil-IRI manual: (a) group screening to identify learners at risk; (b) individualized graded-passage assessment to establish baseline level and profile in each modality (oral, silent, listening); (c) targeted remediation delivered by the school; and (d) post-testing to verify movement across decision levels independent, instructional, frustration [6]. Results were compiled using standard division templates and reported at learner, class, and school levels to inform ECARP-related planning [9].

After the school's internal consolidation, we obtained a de-identified dataset for analysis. Extracted fields included grade level (7–10), sex, age (years), assessment language (English/Filipino), modality (oral, silent, listening), and the official Phil-IRI decision level per modality. We conducted routine data checks—removing duplicate IDs, verifying age ranges, checking grade–age consistency, and scanning cross-tabulations for impossible combinations—and resolved any discrepancies with the reading coordinator before locking the analysis file.

We analyzed the Phil-IRI (Filipino) post-test scores recorded by the school and summarized them by subgroup using counts/percentages, medians (Mdn), and mean ranks (MR) for each demographic category—age bands (11–12, 13–14, 15–16, 17–18), sex (male/female), and grade level (Grades 7–10). Descriptive tables report the number of respondents per group alongside Mdn and their corresponding rank distributions. Given the ordinal nature of the outcome, we used non-parametric tests: Mann–Whitney (Wilcoxon rank-sum) for the two-group sex comparison and Kruskal–Wallis H tests for the multi-group age and grade comparisons. We set α = .05 (two-tailed). For each test, we report the appropriate statistic (U, or W as returned by software, and H with df), exact p-values, and the supporting Mdn/MR to aid interpretation. Post-hoc pairwise procedures (e.g., Dunn with Holm adjustment) were planned only if an omnibus Kruskal–Wallis test reached significance.

3. RESULTS AND DISCUSSION

3.1. Demographic profile of respondents

Table 1 profiles 200 JHS students across four age bands, two sex groups, and four grade levels. Ages cluster in early to mid-adolescence, with 13–14 years comprising half of the cohort (n = 101, 50.5%), followed by 15–16 (n = 58, 29.0%) and 11–12 (n = 37, 18.5%). A small older tail at 17–18 is present (n = 4, 2.0%). Sex is essentially balanced (female = 101, 50.5%; male = 99, 49.5%). Grade levels are evenly represented with 50 learners each in Grades 7–10. This distribution is typical of a large Philippine public junior-high population, and it provides well-balanced strata for the planned rank-based comparisons by sex and grade. The small 17–18 subgroup signals potential over-age status, which warrants cautious interpretation in age-based contrasts and motivates sensitivity checks.

Table 1. Demographic profile of the Students (n = 200).

Demographic Profile		n	%
Age (years)	11-12	37	18.5
	13-14	101	50.5
	15-16	58	29.0
	17-18	4	2.0
	Subtotal	200	100.0
Sex	Male	99	49.5
	Female	101	50.5
	Subtotal	200	100.0
Grade level	Grade 7	50	25.0
	Grade 8	50	25.0
	Grade 9	50	25.0
	Grade 10	50	25.0
	Subtotal	200	100.0

Analytically, the near 1:1 sex ratio reduces the risk of unequal-group confounding and improves the precision of Mann–Whitney comparisons that follow. The equal allocation by grade supports like-for-like contrasts across curricular exposure while keeping statistical power symmetric across groups. By contrast, the sparse older band (17–18) can inflate sampling error and widen uncertainty around age effects. To manage this, we pre-specify two safeguards: reporting exact p values for small-sample contrasts and conducting a sensitivity analysis that either (a) collapses 17–18 into the adjacent 15–16 band or (b) repeats the test excluding 17–18 to verify that any observed pattern is not driven by four cases [25]. These choices are consistent with good practice for nonparametric tests on skewed or sparse strata.

Substantively, the presence of older students in a junior-high cohort may reflect late school entry, repetition, or interrupted schooling. International monitoring consistently links over-age status to weaker reading performance and engagement, partly through cumulative disadvantage and curriculum misalignment [3]. If that pattern holds locally, one would expect the 17–18 subgroup to include a higher share of learners below proficiency in Filipino reading comprehension. The current study tests that expectation using rank-based methods and reports distributional summaries to show whether any apparent gap is systematic or driven by a few extreme values. Regardless of statistical significance, identifying older students who remain below proficiency is programmatically important because remedial needs often differ from those of on-age peers.

The grade-level balance is also informative for interpretation. Equal cell sizes enable clearer comparisons of within-grade variability rather than assuming uniform progress by year level. In reading development, plateaus or non-linear gains across Grades 7–10 are plausible because language and text complexity rise while instructional time and prior skill growth vary [26]. The subsequent Results therefore report Mdn and interquartile ranges by grade, followed by Kruskal–Wallis tests with effect sizes, to distinguish trivial from practically meaningful differences.

These demographics have immediate implications for school-level decision-making. Balanced sex and grade strata allow the school to locate where support is most needed without confounding from uneven enrollment. The small but visible older subgroup argues for routine monitoring systems that flag over-age learners who remain below proficiency, coupled with catch-up supports that are sensitive to both skill level and curriculum demands [6]. In short, the cohort structure is adequate for robust nonparametric subgroup tests, and it already points to actionable priorities that the remainder of the paper elaborates with outcome data.

3.2. Phil-IRI results across demographic groups

Table 2 presents the Phil-IRI post-test scores by demographic group. In this cohort (n = 200), Phil-IRI Filipino post-test scores vary descriptively by age and sex and only modestly by grade. Younger students tend to post higher outcomes: the 11–12 group shows the highest central tendency (Mdn = 26.0, MR = 114.7, n = 37), followed by 13–14 (Mdn = 24.0, MR = 102.3, n = 101) and 15–16 (Mdn = 20.0, MR = 91.6, n = 58), while the small 17–18 subgroup is lowest (Mdn = 15.5, MR = 51.6, n = 4). By sex, females outperform males on both location and ordering (female: Mdn = 26.0, MR = 110.4, n = 101; male: Mdn = 19.0, MR = 90.4, n = 99), indicating that girls are not only more likely to reach higher scores but are also more densely represented near the top of the pooled rank distribution. Grade Mdn are close and non-monotonic (G7: Mdn = 22.5, MR = 96.2, n = 50; G8: Mdn = 24.5, MR = 111.1, n = 50; G9: Mdn = 23.5, MR = 96.2, n = 50; G10: Mdn = 21.5, MR = 98.5, n = 50). Read together, these descriptive signals suggest that time-in-grade alone does not guarantee gains on this screening assessment; rather, cohort composition (e.g., over-age learners) and instruction within grades likely shape outcomes more than the nominal year level [4], [6].

Table 2. Phil-IRI post-test scores by demographic group (n = 200)

Demographic group		n	Median (Mdn)	Mean rank (MR)
Age (years)	11-12	37	26.0	114.7
	13-14	101	24.0	102.3
	15-16	58	20.0	91.6
	17-18	4	15.5	51.6
	Male	101	26.0	110.4
	Female	99	19.0	90.4
	Grade 7	50	22.5	96.2
	Grade 8	50	24.5	111.1
	Grade 9	50	23.5	96.2
	Grade 10	50	21.5	98.5

The female advantage is sizable in both Mdn difference (26.0 vs. 19.0) and rank ordering (MR = 110.4 vs. 90.4). Two complementary explanations from the theory introduced earlier help interpret this pattern. First, under the Simple View of Reading, comprehension depends on the joint strength of word recognition and linguistic comprehension; subgroup gaps emerge when one component lags even as students advance in grade [12], [13]. The distributional tilt toward higher ranks among girls is consistent with stronger average component skills (e.g., more efficient decoding or broader oral/academic language) accumulating into better comprehension. Second, expectancy–value accounts describe a motivational pathway: if one subgroup assigns higher value to reading and expects success, increased persistence and practice follow, which typically widens observed differences on comprehension measures; the alignment of both Mdn and MR in favor of females fits that mechanism [16]. An alternative but compatible reading is ecological: task design, text selection, and discourse norms can differentially elicit engagement. This is where classroom interventions matter. Philippine classroom experiences with structured, gamified literacy tasks and problem-based projects strands reflected in Funa and colleagues' work on engagement-oriented designs offer a practical route to raise perceived value and sustained practice among boys without diluting what already benefits girls. In other words, the gap is not treated as fixed; it is a target for design [27]-[31].

The age gradient (11–12 highest; 17–18 lowest) invites careful interpretation. On one hand, it may capture real skill differences: older students in junior high often include repeaters or those with interrupted schooling whose foundational decoding or language-comprehension weaknesses persist in the absence of targeted intervention [4]. On the other hand, measurement and composition issues may contribute. First, Phil-IRI is a placement and progress-monitoring tool; when cohorts contain many students clustered in the instructional band, Mdn can be "sticky," and small older subgroups (n = 4 for 17–18) inflate sampling error. Second, text familiarity and the language of assessment (Filipino) may differentially advantage students whose recent reading exposure has been dominantly in that language; older learners who shifted reading time to subject-specific English texts may not translate that exposure into gains on a Filipino comprehension screen. These alternatives do not undermine the signal; they refine it. The programmatic upshot is the same: older learners remaining in junior high are a priority group for diagnostics and short, frequent, profile-matched supports (e.g., decoding fluency sprints versus oral language scaffolds), with documented movement across Phil-IRI decision levels. Where schools layer digital practice (e.g., guided reading or vocabulary rehearsal), the line from Funa's work on responsible, teacher-directed technology use is relevant: transparency, human oversight, and equity safeguards ensure that technology augments, rather than replaces, explicit instruction [6], [29]-[31].

The near-flat, non-monotonic grade pattern (G8 slightly higher; G7 ≈ G9; G10 modestly lower) argues against planning from grade averages. Three non-exclusive mechanisms are plausible and consistent with the frameworks used in the Introduction. First, cohort-mix: within-grade variability can be large (SVR predicts mixed profiles—"poor decoders," "poor comprehenders," and "generally weak readers"—coexisting in a single section); grade Mdn obscure these subtypes [12], [13]. Second, instructional alignment: when the linguistic demands of texts rise faster than students' component skills, comprehension stalls; Mdn remain flat even if word recognition improves because linguistic comprehension becomes the bottleneck [13]. Third, engagement dynamics: expectancy–value predicts that perceived cost and low value can offset gains in skill; students may read "just enough" to get by unless tasks make value visible (relevance) and success probable (scaffolded challenge) [16]. These mechanisms are not speculative add-ons; they are testable drivers that inform how Phil-IRI data should be used in the school improvement cycle.

Two counter-arguments deserve attention. One might argue that the female–male difference could simply reflect test-format familiarity (e.g., girls reading more outside school) rather than true skill differences. Yet the observed pattern involves both central tendency (Mdn) and ordering (MR) across the full distribution, which is exactly what one would expect from accumulated, domain-general reading experiences that strengthen component skills and build strategic knowledge [4]. Another counter-argument is that the age gradient is an artifact of tiny n in the 17–18 band. We agree that uncertainty is high for that cell; accordingly, the inferential section pre-specifies sensitivity checks that pool 17–18 with 15–16 or temporarily exclude 17–18 to verify that any “older-learner lag” is not driven by four cases. The point is not to over-read a small cell; it is to set rules that protect conclusions while still responding to a recognizable, programmatically important subgroup [4].

These interpretations lead directly to action. To ground these actions in evidence, engagement systems such as digital badges can increase persistence and make progress visible in secondary classrooms, which aligns with Phil-IRI’s diagnose–match–monitor cycle and may help sustain daily reading practice [32]. Structured practice models from clinical education similarly emphasize feedback loops, supervised rotations, and authentic tasks; in literacy terms, this translates into short, frequent, level-appropriate reading with immediate feedback and scheduled reviews using Phil-IRI decision levels [33]. Finally, culturally adapted measurement work underscores the value of interpreting learner signals with context-sensitive tools; applying this lens to Phil-IRI helps teams distinguish decoding versus language-comprehension constraints and respond proportionately within grades [34]. These principles are consistent with international and local evidence on motivation, progress monitoring, and responsive tiered supports in secondary literacy [35]–[38].

For sex differences, the goal is boy-responsive engagement that preserves what already works for girls: rotate interest-driven texts (sports, STEM curiosities, adventure); embed structured practice with immediate feedback; and make progress visible (reading logs, brief retells, and teacher conferences). For age, build a diagnose–match–monitor loop: (a) use Phil-IRI procedures to classify likely profiles (decoding vs. language-comprehension constraints), (b) assign short, frequent interventions (10–15 minutes/day) aligned to that profile, and (c) track transition probabilities between Phil-IRI decision levels each quarter. For grade, shift from year-level averages to within-grade tiering: use class distributions to assign small-group instruction, integrate listening comprehension for classes trending “poor comprehender,” and schedule fluency sprints where “poor decoder” signs dominate. Where schools pilot AI-supported practice (e.g., adaptive vocabulary review or fluency reading with speech feedback), apply responsible-use guardrails emphasized in Philippine policy scholarship to ensure alignment with DepEd guidance and to avoid widening gaps a theme underscored in Funa’s work on integrating technology prudently into instruction [6], [29]-[31].

3.3 Group differences in Phil-IRI scores by age, sex, and grade

Below we discuss each demographic variable in turn. For each, we restate the pattern in the data, interpret what it implies about students’ reading performance, situate the finding within the study’s theoretical frame (Simple View of Reading and expectancy–value), and draw concrete implications for school practice.

Table 3 presents the difference between the respondents’ demographic profile and their Phil-IRI scores. The rank-based comparison across the four age bands did not reach statistical significance (Kruskal–Wallis: $H = 6.58$, $df = 3$, $p = .087$). Descriptively, however, scores trended downward with age: 11–12 years had the highest Mdn (26.0; MR = 114.7), followed by 13–14 (24.0; 102.3) and 15–16 (20.0; 91.6), while the small 17–18 group ($n = 4$) showed the lowest center (15.5; 51.6). This suggests that the oldest students—likely including late entrants or repeaters—cluster toward the lower end of the distribution, but the combination of overlapping Mdn and a very small n in the oldest band limits statistical power to detect a reliable omnibus difference [4].

Table 3. Difference between the respondents’ demographic profile and their Phil-IRI scores

Contrast	Test	Statistic (df)	p-value	α	Decision	Interpretation
Age (11–12, 13–14, 15–16, 17–18)	Kruskal–Wallis	$H = 6.58$ (df = 3)	.087	.05	Fail to reject H_0	No significant difference by age
Sex (female vs male)	Mann–Whitney U	$U = 5{,}999.4$	< .001	.05	Reject H_0	Females score higher than males
Grade level (G7, G8, G9, G10)	Kruskal–Wallis	$H = 2.30$ (df = 3)	.513	.05	Fail to reject H_0	No significant difference by grade

Interpreted through the SVR, the pattern is consistent with heterogeneous profiles in word recognition and language comprehension within each age band: some learners accrue years of schooling without sufficient gains in one or both components, so within-age variance outweighs between-age differences [12], [13]. From an expectancy–value lens, over-age students who have struggled in prior years may show lower expectancy (“Can I do this?”) and higher perceived costs, dampening engagement and slowing progress even as chronological age rises [16].

Age alone is a poor proxy for reading proficiency in this cohort. Schools should (1) flag over-age learners for closer diagnostic review, (2) deliver targeted decoding and language-comprehension supports based on Phil-IRI profiles rather than age, and (3) monitor movement across Phil-IRI decision levels during the year [6]. When analyzing future cohorts, consider collapsing sparse age bands or using an over-age indicator (e.g., ≥2 years over typical grade age) to improve interpretability and power [4].

The sex comparison was significant (Mann–Whitney U = 5,999.4, p < .001), indicating higher Phil-IRI scores among females. Descriptively, females (n = 101) had an Mdn of 26.0 and a MR around 110, while males (n = 99) had an Mdn of 19.0 and a MR around 90. The separation in both Mdn and MR shows that girls are not only higher at the center but are consistently distributed toward the top of the overall ranking (see also [4], on a 35-point female advantage in reading in PISA 2022 for the Philippines).

The result aligns with the broader literature and is interpretable within our theoretical frame. Under SVR, subgroup gaps can reflect imbalances in component skills; combined with expectancy–value, girls' generally higher reading enjoyment and task value translate into more practice and persistence, reinforcing advantages in vocabulary, fluency, and comprehension [12], [13], [16]. In short, the female advantage observed here is both statistically reliable and theoretically coherent.

Maintain practices that are working for girls while closing the gap for boys. Priorities include (1) boy-responsive engagement—high-interest, choice-rich texts; short goal-cycles with visible progress; structured accountability for reading—an approach consistent with classroom evidence on gamified/engagement-focused designs [30], (2) explicit fluency and vocabulary routines (repeated reading, morphology/word-study) integrated into content classes [3932], and (3) tight progress monitoring using Phil-IRI levels to verify movement from frustration/instructional to independent [6]. Teachers can review sex-disaggregated Phil-IRI data in PLCs and adjust materials, pacing, and feedback loops accordingly; structured, problem-based tasks have also shown positive effects on secondary achievement, supporting systematic practice [29].

Across Grades 7–10, the omnibus test was not significant (Kruskal–Wallis: H = 2.30, df = 3, p = .513). Mdn were close and non-monotonic—Grade 8 had the highest center (24.5; MR = 111.1), Grades 7 and 9 were similar (22.5 and 23.5; MR = 96.2), and Grade 10 was slightly lower (21.5; 98.5). Thus, simple promotion by year level does not guarantee higher Phil-IRI performance. Within SVR, this is expected when classrooms contain varied mixtures of "poor decoder" and "poor comprehender" profiles; grade averages can mask sizable within-grade needs [12], [13]. Differences in exposure, curriculum implementation, teacher practices, or absenteeism can also offset the effect of an additional year of schooling, yielding the observed flat profile [16], [4033].

Plan supports within grades, not for grades as a whole. Practical steps are to (1) allocate tiered interventions (independent/instructional/frustration) inside each grade, (2) use common grade-level assessments and shared interventions so teachers can compare like with like, and (3) set within-grade improvement targets (e.g., reduce the proportion at frustration by X% each quarter) verified by Phil-IRI post-checks [6]. Where Grade 8 appears relatively stronger, examine practices that can be transferred to Grades 7, 9, and 10.

In this cohort, sex is the only demographic variable with a statistically reliable association with Phil-IRI scores; age and grade level are not. Instructional decisions should therefore prioritize sex-responsive engagement and within-grade, profile-based remediation, using Phil-IRI's decision levels to monitor whether students move toward independence over the school year [6].

4. CONCLUSION

Aligned with the study objectives, three points stand out. First, the cohort composition was suitable for subgroup analysis: sex was nearly balanced, grade levels were evenly represented, and ages clustered in early to mid-adolescence, with a small older subgroup that reduces power for age contrasts. Second, the descriptive profile suggested a female advantage in reading, modest and non-monotonic differences across grades, and lower centers among the oldest age band. Third, inferential tests confirmed a reliable difference by sex, favoring females, while finding no evidence of differences by age or by grade level. Overall, Phil-IRI outcomes in this school vary more within grades than between grades, and demographic markers other than sex explain little of the observed variation.

Given the study's patterns, schools may organize support within grades rather than by grade level, using Phil-IRI decision levels to group learners and align instruction with identified needs, with termly reviews to verify movement toward independence. Because sex was the only demographic factor associated with Phil-IRI performance, classrooms may introduce modest, boy-responsive engagement routines that increase time on task and make progress visible while maintaining practices that already benefit girls. As age and grade showed no significant differences, resources may be allocated without presuming year-level effects; however, teachers may monitor the small older subgroup that appeared weaker in the descriptives and provide case-by-case assistance as indicated. These adjustments may be tracked through sex-disaggregated summaries and integrated into regular school planning.

REFERENCES

[1] T. Kheang, P. Chin, and S. Em, "Reading motivation to promote students' reading comprehension: A review," *International Journal of Languages and Culture*, vol. 4, no. 1, pp. 1–8, 2024, doi: 10.51483/IJLC.4.1.2024.1-8.

[2] K. Mastrothanasis, M. Kladaki, and A. Andreou, "A systematic review and meta-analysis of the Readers' Theatre impact on the development of reading skills," *International Journal of Educational Research Open*, vol. 4, Art. no. 100243, 2023, doi: 10.1016/j.ijedro.2023.100243.

[3] OECD, *PISA 2018 Results (Volume I): What Students Know and Can Do*. Paris, France: OECD Publishing, 2019, doi: 10.1787/5f07c754-en.

[4] OECD, *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. Paris, France: OECD Publishing, 2023, doi: 10.1787/83b2b81c-en.

[5] UNESCO Institute for Statistics, "UIS education data browser (minimum proficiency in reading)," Montreal, Canada: UIS, Sept. 2023 release.

[6] DepEd (Philippines), "Policy guidelines on the Administration of the Revised Philippine Informal Reading Inventory (Phil-IRI)," Mar. 26, 2018. [Online]. Accessed: Aug. 15, 2025.

[7] DepEd (Philippines), *Philippine Informal Reading Inventory (Phil-IRI) Manual*, 2018 ed. Manila, Philippines: DepEd, 2018. [Online]. Accessed: Aug. 20, 2025.

[8] DepEd (Philippines), "DO 70, s. 2011 — guidelines on the utilization of funds for the every child a reader program (ECARP)," 2011. [Online]. Accessed: Aug. 12, 2025.

[9] DepEd (Philippines), "DM 143, s. 2012 — Administration of the philippine informal reading inventory (Phil-IRI) Assessment Tools," 2012. [Online]. Accessed: Aug. 21, 2025.

[10] DepEd Region VIII (Eastern Visayas), "RM — Administration of the Phil-IRI for Key Stages 2 and 3, SY 2024–2025," 2024. [Online]. Accessed: Aug. 22, 2025.

[11] Schools Division of Nueva Vizcaya, "DM-CT-2024-447 — Administration of the Philippine informal reading inventory (phil-iri) for key stages 2 and 3, SY 2024–2025," Nov. 29, 2024.

[12] P. B. Gough and W. E. Tunmer, "Decoding, reading, and reading disability," *Remedial and Special Education*, vol. 7, no. 1, pp. 6–10, 1986, doi: 10.1177/074193258600700104.

[13] H. W. Catts, "The simple view of reading: Advancements and false impressions," *Remedial and Special Education*, vol. 39, no. 5, pp. 317–323, 2018, doi: 10.1177/0741932518767563.

[14] H. S. Scarborough, "Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice," in *Handbook of Early Literacy Research*, vol. 1, S. B. Neuman and D. K. Dickinson, Eds. New York, NY, USA: Guilford, 2001, pp. 97–110.

[15] C. A. Perfetti and J. Stafura, "Word knowledge in a theory of reading comprehension," *Scientific Studies of Reading*, vol. 18, no. 1, pp. 22–37, 2014, doi: 10.1080/10888438.2013.827687.

[16] J. S. Eccles and A. Wigfield, "Motivational beliefs, values, and goals," *Annual Review of Psychology*, vol. 53, pp. 109–132, 2002, doi: 10.1146/annurev.psych.53.100901.135153.

[17] M. van Hek, C. Buchmann, and G. Kraaykamp, "Educational systems and gender differences in reading: A comparative multilevel analysis," *European Sociological Review*, vol. 35, no. 2, pp. 169–186, 2019, doi: 10.1093/esr/jcy054.

[18] C. L. Pagal, J. K. A. Mirafuentes, and Q. C. Ypanto, "School age gender gap in reading comprehension," *Journal of Asian Development*, vol. 3, no. 2, pp. 78–102, 2017, doi: 10.5296/jad.v3i2.11079.

[19] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 5th ed. Thousand Oaks, CA, USA: SAGE, 2018.

[20] K. A. Levin, "Study design III: Cross-sectional studies," *Evidence-Based Dentistry*, vol. 7, pp. 24–25, 2006, doi: 10.1038/sj.ebd.6400375.

[21] M. S. Setia, "Methodology series module 3: Cross-sectional studies," *Indian Journal of Dermatology*, vol. 61, no. 3, pp. 261–264, 2016, doi: 10.4103/0019-5154.182410.

[22] F. N. Kerlinger and H. B. Lee, *Foundations of Behavioral Research*, 4th ed. Fort Worth, TX, USA: Harcourt/Wadsworth (Thomson Learning), 2000.

[23] N. J. Salkind, Ed., *Encyclopedia of Research Design*. Thousand Oaks, CA, USA: SAGE, 2010.

[24] E. von Elm *et al*., "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies," *PLoS Medicine*, vol. 4, no. 10, e296, 2007, doi: 10.1371/journal.pmed.0040296.

[25] J. H. McDonald, *Handbook of Biological Statistics*, 3rd ed. Baltimore, MD, USA: Sparky House Publishing, 2014.

[26] A. N. Spichtig, J. P. Pascoe, K. M. Gehsmann, F. Gu, and J. D. Ferrara, "The interaction of silent reading rate, academic vocabulary, and comprehension among students in grades 2–12," *Reading Research Quarterly*, vol. 57, no. 3, pp. 1003–1019, 2022, doi: 10.1002/rrq.457.

[27] A. A. Funa and R. A. E. Gabay, "Policy guidelines and recommendations on AI use in teaching and learning: A meta-synthesis study," *Social Sciences & Humanities Open*, vol. 11, art. 101221, 2025, doi: 10.1016/j.ssaho.2024.101221.

[28] A. A. Funa, R. A. E. Gabay, and J. D. Ricafort, "Gamification, gameplay, and genetics learning: A meta-analysis," *Jurnal Pendidikan IPA Indonesia*, vol. 10, no. 4, pp. 462–473, 2021, doi: 10.15294/jpii.v10i4.32143.

[29] A. A. Funa and R. A. E. Gabay, "Bridging the gap among knowledge, attitudes, and behaviors toward sustainable development through I-STEM-PBL-ESD," *International Journal on Studies in Education (IJonSE)*, vol. 7, no. 2, pp. 288–303, Apr. 2025, doi: 10.46328/ijonse.317.

[30] A. A. Funa and J. D. Ricafort, "Validation and effectiveness of computer-based gamified instructional material for teaching introductory physics," *International Journal of Sciences: Basic and Applied Research*, vol. 47, no. 2, pp. 168–180, 2019.

[31] A. A. Funa, R. A. E. Gabay, K. L. Esdicul, and M. S. Prudente, "Secondary teachers' and students' perceptions of distance education in science: Focus on learner-centered, action-oriented, and transformative learning," Dalat Univ. J. Sci., vol. 13, no. 3, pp. 156–181, 2023, doi: 10.37569/DalatUniversity.13.3.1108.A. A. Funa, "Digital badges as

rewards in science education: Students' perceptions and experiences," Dalat University Journal of Science, vol. 14, no. 2, pp. 107–127, 2024. doi: 10.37569/DalatUniversity.14.2.1212(2024).
[32] A. A. Funa, "Exploring clinical practice education for student midwives: Hospital and community opportunities and challenges," Midwifery, vol. 139, p. 104186, 2024. doi: 10.1016/j.midw.2024.104186.
[33] A. A. Funa and R. A. E. Gabay, "Cultural adaptation of predictive processing measures: Insights from the Prediction-Related Experiences Questionnaire (PRE-Q) for the Philippine context," Asian Journal of Psychiatry, vol. 112, p. 104681, 2025. doi: 10.1016/j.ajp.2025.104681.
[34] J. R. Toste, L. Didion, P. Peng, M. J. Filderman, and A. M. McClelland, "A Meta-Analytic Review of the Relations Between Motivation and Reading Achievement for K–12 Students," Review of Educational Research, vol. 90, no. 3, pp. 420–456, 2020. doi: 10.3102/0034654320919352.
[35] N. K. Scammacca, G. Roberts, S. Vaughn, and K. K. Stuebing, "A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011," Journal of Learning Disabilities, vol. 48, no. 4, pp. 369–390, 2015. doi: 10.1177/0022219413504995.
[36] D. Fuchs, L. S. Fuchs, and D. L. Compton, "Smart RTI: A next-generation approach to multilevel prevention," Exceptional Children, vol. 78, no. 3, pp. 263–279, 2012. doi: 10.1177/001440291207800301.
[37] What Works Clearinghouse, Providing Reading Interventions for Students in Grades 4–9 (WWC 2022007), Washington, DC: U.S. Department of education, institute of education sciences, 2022. URL: ies.ed.gov/ncee/wwc/practiceguide/29.
[38] J. M. Ramallosa, A. A. Funa, A. T. Geron, R. T. Ibardaloza, and M. S. Prudente, "Meta-analysis on the effectiveness of argument-based learning on students' conceptual understanding," in *Proc. 13th Int. Conf. on E-Education, E-Business, E-Management and E-Learning (IC4E)*, Tokyo, Japan, Jan. 14–17, 2022, pp. 315–323, doi: 10.1145/3514262.3514305.
[39] N. L. Mediana, Jr., A. A. Funa, and R. V. Dio, "Effectiveness of inquiry-based learning (IbL) on improving students' conceptual understanding in science and mathematics: A meta-analysis," *International Journal of Education in Mathematics, Science and Technology*, vol. 13, no. 2, pp. 532–552, 2025, doi: 10.46328/ijemst.4769.