Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Improving Learner Performance Through Score Based Analysis on a Gamified AI Driven Cybersecurity Learning Platform

Jean Emmanuel Ntsama
Modibbo Adama University
Gregory Maksha Wajiga
Modibbo Adama University
Etemi Joshua Garba
Modibbo Adama University
Franklin Tchakounté
University of Ngaoundéré
Derrick Méthode Bagaza
dmbagaza@univ-bangui.org
University of Bangui
Faissal Ahmadou
University of Ngaoundéré

Research Article

Keywords: cybersecurity education, AI-driven learning, gamification, personalized learning, learner performance

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11139062/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Improving Learner Performance Through Score Based Analysis on a Gamified AI Driven Cybersecurity Learning Platform

Jean Emmanuel Ntsama[1,3], Gregory Maksha Wajiga[1], Etemi Joshua Garba[1], Franklin Tchakounté[2,3], Derrick Méthode Bagaza[4*], Faissal Ahmadou[2,3]

[1]Department of Computer Science, Faculty of Computing, Modibbo Adama University, Yola, Nigeria.
[2]Department of Mathematics and Computer Science, Faculty of Science, University of Ngaoundéré, Cameroon.
[3]Cybersecurity with Computational and Artificial Intelligence Group (CyComAI), Cameroon.
[4*]Department of Computer Science, Higher Institute of Technology, University of Bangui, Central African Republic.

*Corresponding author(s). E-mail(s): dmbagaza@univ-bangui.org;
Contributing authors: ntsama.jemma@gmail.com; gwajiga@mau.edu.ng; e.j.garba@mau.edu.ng; tchafros@gmail.com; faissalahmadou@gmail.com;

Abstract

This study examines whether structured engagement with instructional content improves learner performance on Cyber Security Academy, an AI-driven, gamified cybersecurity learning platform. Activity logs covering 1,500 completed assessments from users in ten Anglophone and Francophone countries were analyzed by comparing learners who consulted the lessons and completed the full learning pathway (n = 900) with those who engaged only partially with the platform (n = 600). Group differences were tested with Welch's t-test, the Mann Whitney U test, one-way ANOVA, Spearman correlations, and Cohen's d as a measure of effect size. Learners who read the lessons before assessment obtained a mean score of 71.27% versus 60.41% for non-readers ($\mathbf{\Delta} = +10.86$ points; t(1192) = 12.21; p ¡ 0.001; d = 0.650; 95% CI [9.12, 12.60]), committed fewer errors (2.50 vs. 4.85; d = −1.069), and responded faster (9.85 vs. 12.69 minutes; d = −0.807), indicating greater fluency of cognitive processing. The effect decreased

monotonically with age (d = 1.205 for learners aged 13 and under; d = 0.387 for those 22 and over) and with digital literacy (d = 1.434 for beginners; d = 0.524 for experts), while remaining positive in all ten countries. These findings indicate that structured, AI-personalized, and gamified learning environments substantially enhance cybersecurity learning outcomes, particularly for younger and less experienced users.

Keywords: cybersecurity education, AI-driven learning, gamification, personalized learning, learner performance

1 Introduction

Cybersecurity education has progressively shifted from static, lecture-based instruction toward interactive and adaptive environments capable of simulating real-world cyber threats. Traditional approaches often struggle to sustain learner engagement and to develop the practical problem-solving skills demanded by contemporary security contexts. Gamified learning platforms have emerged as effective alternatives, integrating points, badges, leaderboards, challenges, and scenario-based activities to enhance motivation, participation, and knowledge retention (Hamari et al., 2016; Krombholz et al., 2015). In parallel, advances in artificial intelligence (AI) enable personalized learning: adaptive systems can adjust content difficulty, recommend individualized pathways, and deliver real-time feedback according to each learner's performance and skill gaps (Ras et al., 2023). The combination of gamification and AI personalization therefore represents a promising approach for improving both engagement and learning effectiveness in cybersecurity training.

Despite the growing adoption of these technologies, one question remains insufficiently explored: do learners actually perform better when they follow a structured, complete learning pathway rather than using a platform in a fragmented manner? Many learners interact selectively with online platforms, skipping lessons, exercises, or remediation activities, which may limit the pedagogical effectiveness of adaptive and gamified systems. Empirical evidence on the relationship between structured engagement and learning outcomes is therefore needed.

This study investigates the educational effectiveness of Cyber Security Academy, an AI-driven and gamified cybersecurity learning environment combining adaptive learning sequences, scenario-based challenges, and personalized feedback. It evaluates whether learners who complete the full instructional pathway—lessons, exercises, assessments, and remediation—outperform those who engage only partially. Performance is analyzed through assessment scores, error counts, and response times, with demographic and digital-literacy characteristics used to contextualize group comparability. The main contribution of this paper is empirical evidence on the pedagogical value of combining AI personalization with gamification in cybersecurity education, with particular attention to how the benefit of structured engagement varies with age and digital literacy.

The remainder of the paper is organized as follows. Section 2 reviews related work; Section 3 presents the methodology; Section 4 reports the analyses and results; Section 5 discusses the findings and their limitations; Section 6 concludes.

2 Related Work

A growing body of research indicates that gamified and AI-enhanced platforms can significantly improve learner performance in cybersecurity education. When gamification elements: points, badges, challenges, narratives are combined with adaptive AI, learners exhibit higher engagement, better knowledge retention, and stronger practical-skill development than in traditional settings (Hamari et al., 2016; Peters & Bulterman, 2020). Systematic reviews of gamification in cybersecurity education converge on two findings: gamification increases motivation and completion rates, and AI-driven personalization improves learning efficiency by adjusting difficulty and content in real time (Peters & Bulterman, 2020; Ras et al., 2023). These reviews also note that most studies target university courses or enterprise security awareness, leaving broader, mixed-level user populations less explored.

Research on AI-supported gamified laboratories and personalized cybersecurity platforms further shows that learners who receive tailored, scenario-based challenges with immediate feedback perform better on assessments and report greater satisfaction than those exposed to one-size-fits-all materials (Ras et al., 2023; Shute & Ventura, 2013). From a theoretical standpoint, these gains are consistent with cognitive load theory, which holds that structured instructional guidance reduces extraneous load and supports schema acquisition (Sweller, 1988; Sweller et al., 2011), and with the expertise reversal effect, whereby guidance benefits novices more than experts (Kalyuga, 2007; Sweller & Chandler, 1994). Recent studies suggest that gamified and AI-supported cybersecurity learning improves student engagement and achievement, although most findings remain limited by methodological weakness (Click Armor, 2025; UjPR Authors, 2025; Ntsama et al., 2026). The present study builds on this background by examining, on a large multi-country sample, how completing a full AI-gamified learning sequence affects assessment scores, errors, and response times compared with partial engagement.

Table 1: Overview of related studies on gamified and AI-driven cybersecurity learning

Author(s)	Year	Title	Focus Area	Type of Study
Hamari, J. et al.	2016	Gamification and gamified designs in education: A literature review	General gamification in education, including technical domains	Review and synthesis of empirical studies
Click Armor	2025	Gamified Learning Platform for cybersecurity awareness	Engagement and retention via gamified security awareness labs	Descriptive case study / platform report
HRMARS review team	2025	Cybersecurity Education Using Gamification: Systematic Literature Review	Effectiveness and adoption of gamification in cybersecurity ed	Systematic Literature Review
AIS EL (gen. AI gamified labs)	2025	Generative AI in Gamified Cybersecurity Education	AI supported gamified labs in classroom cybersecurity courses	Empirical study and design research
PeerJ authors	2025	Unveiling personalized and gamification based cybersecurity risks (application in finance)	Personalized, gamification based risk mitigation training	Empirical study in applied cybersecurity
IJrPR authors	2025	Cyber Learning Gamification Platform (AI adaptive, gamified cybersecurity)	AI driven, gamified cybersecurity platform with adaptive paths	Platform design and impact evaluation
Hoxhunt	2025	Does Gamified Cyber Security Training Actually Work? (blog based synthesis)	Engagement, completion, and behavioral change from gamification	Applied industry report / meta synthesis
Click Armor	2025	Security Awareness Labs	Gamified labs improve individual engagement and security awareness retention.	Descriptive Case Study
AIS EL	2025	Generative AI Labs	Generative AI adjusts task difficulty in real time within classroom spaces.	Design Research
IJrPR Authors	2025	Cyber Learning Platform	AI-driven adaptive paths maximize individual student completion metrics.	Impact Evaluation
Ntsama et al.	2026	Cyber Security Academy	Structured pathway raised scores by 10.86 % (M=71.27% vs. 60.41%, 95% CI [9.12, 12.60]), pj0.001, d = 0.650).	Quasi-Experimental N=1500

3 Methodology

3.1 Research Design

The study adopted a quasi-experimental, observational design with two naturally occurring (non-randomized) groups. The independent variable was the completion status of the CyberSecurity Academy learning pathway, operationalized through prior lesson reading (has_read_lessons: HRL): Group 1 (structured learners, $n_1 = 900$) consulted the lessons and completed the prescribed sequence of modules, quizzes, and practical challenges; Group 2 (partial users, $n_2 = 600$) had recorded scores but did not complete the recommended pathway. The dependent variables were the platform's assessment score (0 to 100), the number of errors committed (0 to 15), self-reported mastery of digital tools or digital literacy, and the assessment response time in minutes.

3.2 Data source and nature

Data were drawn from the activity logs of the platform (cybersecurity-edu.cm) collected through Firebase Firestore, covering January 2024 to March 2026. After consolidation and cleaning, the dataset comprised 1,500 completed assessments from users in ten countries: Cameroon (442), Côte d'Ivoire (214), France (172), Nigeria (169), DR Congo (154), Burkina Faso (73), Benin (72), Canada (70), Gabon (68), and Togo (66). Only anonymized records for which consent to aggregate use had been granted at registration were accessed; no intervention beyond the users' natural platform behavior occurred.. Data were collected through Firebase Firestore logs, then consolidated and cleaned before analysis. A representative sample of data collected is presented in the table below;

Table 2: Reprentative sample of collected data

user_id	age_group	sex	MDT	Country	Has_Read (HRL)	score	errors	Time (min)
USR01117	22- more	F	I	Cameroon	0	54	7	18,8
USR01369	22- more	M	B	France	0	40	10	5,0
USR00423	14-21	F	B	RD Congo	1	55	4	28,5
USR00414	14-21	F	I	Nigeria	1	88	3	22,4
USR00452	14-21	F	I	Togo	0	53	4	26,5
USR00862	22- more	F	E	Nigeria	1	80	2	16,6
USR01064	22- more	F	I	Cameroon	0	64	2	24,9
USR00742	14-21	M	B	Canada	0	51	7	7,9
USR01273	22- more	M	B	Benin	1	73	4	30,8
USR00260	Under-13	M	I	Gabon	0	54	5	9,6

Beginner (B), Intermediate (I), and Expert (E); HRL: Has Read Lessons or has read (prior reading)

The sample was characterized through descriptive statistics into Score, Sex proportions, Age distribution, Self-reported mastery of digital tools or digital literacy, country, and time taken to show competency.

3.3 Participants

Description of the sample in terms of three age groups and sex: Female, and Male. The reading status variable was distributed as 60% for the group that had read the lessons

following the learning process n1 = 900 and 40% for the group that had not read the lessons n2 = 600), reflecting the proportions observed in the real activity logs.

All participants were regular users who had registered via the platform's web or mobile interface and had engaged with at least one module that generated a score. Participation was non-experimental and non-interventional; no manipulation of conditions occurred beyond the users' natural behavior on the platform.

Table 3: Distribution of the sample by age-group and sex

Age-group/Sex	Female (F)	Male (M)	Total
Age ¡= 13	100 (6.7%)	200 (13.3%)	300 (20,0%)
14 ¡= Age¡= 21	150 (10.0%)	350 (23.3%)	500 (33,3%)
22¡= Age	225 (15.0%)	475 (31.7%)	700 (46,7%)
TOTAL	475 (31.7%)	1025 (68.3%)	1 500 (100%)

The sample (N = 1,500) included 475 females (31.7%) and 1,025 males (68.3%), distributed across three age groups: 13 years and under (300; 20.0%), 14–21 years (500; 33.3%), and 22 years and over (700; 46.7%). Self-reported mastery of digital tools (MDT) was recorded at registration on a three-point scale: beginner (379), intermediate (682), expert (439), and used for descriptive and moderation analyses. The two groups showed broadly similar sex and age structures, with no extreme skewing threatening the validity of the comparison; the slightly higher proportion of adults among partial users is consistent with the more fragmented time availability reported for adult online learners (Hamari et al., 2016).

Table 4: Distribution of the sample by country

Country	Total	Country	Total
Cameroon	442	Burkina Faso	73
Côte d'Ivoire	214	Bénin	72
France	172	Canada	70
Nigeria	169	Gabon	68
RD Congo	154	Togo	66
TOTAL1	1151	TOTAL2	349

3.4 Measures

The primary outcome was the platform's composite score (0–100), computed uniformly for all users as a weighted average of quiz accuracy, challenge completion, and timeliness of responses. Secondary outcomes were the number of errors and the response time (time elapsed between the first and last submitted answers). Contextual variables included sex, age group, digital literacy (MDT), country, and platform stage.

The primary outcome measure was the platform's score at each level, stage and final score, a normalized metric ranging from 0 to 100, computed as a weighted average of: Quiz accuracy (number of correct items divided by total items); challenge completion (percentage of active challenges completed successfully); timeliness of responses (a penalty for late or skipped items, where applicable).

The scoring algorithm was applied uniformly across all users, producing a single composite score that reflects overall performance in the course.

3.5 Study variables

Demographic and Contextual Variables observe were: Sex a Binary self-reported item (female/male); Age as Self-reported age in years, later categorized into three ranges for descriptive analysis; Mastery of digital tools: Self-reported 3-point Likert scale (beginner (B), intermediate (I) and expert (E)), collected at registration to inform the platform's initial difficulty calibration.

The main independent variable was prior reading of lessons: has read_lessons (HRL). The dependent variables were evaluation score (score) and number of errors: nb_of_errors (NOE). Other variables included completion time, age group, platform stage, sex, digital skill level, geographic location, and country (Cohen, J., 1988).

Table 5: Study variables

Variable	Nature	Description	Value
has_read_lessons or has_read (HRL)	Binairy / IV	Principal indépendant Variable	0 = no · 1 = Yes
Score	Continious / DV	Score evaluation (%)	0 to 100
Nb_of_errors (NOE)	Discrete / DV	Number of errors made	0 to 15
completion_min_time	Continious	Time taken (minutes)	5 to 60
age_group	Category	Age interval of the target	Age ¡= 13 / 14¡= Age ¡= 21 / 22¡= Age
Stage	Category	Stage in plateform	Stage 1 / 2 / 3
Sex	Binairy	Sex is declare	F / M
MDT	Ordinal	Mastery of Digital Tools or Digital literacy	Beginer (B) / Intermediate (I) / Expert (E)
Country	Category	Learner country	10 modalities
localization	Binary	Geographical zone	Urban / Rural

MDT: Mastery of Digital Tools HRL: has read lessons or prior reading NOE: Number of errors

3.6 Research hypotheses

The null hypothesis H_0 stated that there would be no significant difference in performance between learners who consulted the pedagogical resources and those who did not. Conversely, the alternative hypothesis H_1 stated that learners who consulted the pedagogical resources and followed the structured learning pathway would obtain significantly higher scores and commit significantly fewer errors than learners who used the platform only partially. This effect was expected to be stronger among children

and adolescents than among adults, in line with educational interpretations of effect size and learner heterogeneity (Cohen, 1988).

Five operational sub-hypotheses were derived from H_1:

- H_1a: The effect of prior lesson reading is stronger among learners aged 13 years and under.
- H_1b: Among adult learners aged 22 years and over, the effect of prior lesson reading is moderated by digital literacy level (MDT).
- H_1c: Learners who consult the lessons obtain higher scores regardless of sex.
- H_1d: Learners who read the lessons before assessment respond more quickly during assessment.
- H_1e: Learner scores vary according to the country in which the platform is used.

Statistical methods

Five complementary methods were used: Welch's t-test for comparing means of independent groups with unequal variances (Welch, 1947), Cohen's d for effect sizes (Cohen, 1988; Lakens, 2013), the Mann–Whitney U test as a non-parametric robustness check, one-way ANOVA for multi-group comparisons, and Spearman correlations for associations between variables. All tests used $\alpha = .05$.

The null hypothesis H_0 stated no performance difference between readers and non-readers. The alternative hypothesis H_1 stated that learners who consulted the lessons within the structured pathway would obtain higher scores and commit fewer errors. Five sub-hypotheses were tested: (H_1a) the reading effect is stronger among learners aged 13 and under; (H_1b) among learners, the effect is moderated by digital literacy; (H_1c) the effect holds for both sexes; (H_1d) readers respond faster during assessment; (H_1e) the effect is stable across countries.

Data were collected from the platform's internal analytics dashboard over a six-month window. The research team accessed only anonymized user records for which consent to use aggregate data had been granted at registration. For each user, the following information was extracted:

- Group membership (partial vs. complete).
- Final score.
- Sex, age, and self-reported mastery of digital tools.

No additional external tests or interventions were administered. The platform integrated courses, exercises, and assessment system served as the sole outcome measure and remediation. The study protocol was approved by the relevant institutional ethics committee (see Declarations), and informed consent to the use of anonymized data for aggregate research was obtained from all participants at registration, prior to data collection.

4 Results

The data analysis followed a three-step approach. The main comparison focuses on assessment scores and the number of errors between the 900 learners who read the

lessons and the 600 who did not. We apply Welch's t-test, the most appropriate method for comparing two means from independent groups with potentially unequal variances.

Descriptive comparison

Readers scored on average 10.86 points higher than non-readers and committed 2.35 fewer errors (Table 6). In both groups the mean and median scores nearly coincide, and skewness is close to zero with slightly negative kurtosis, indicating approximately symmetric distributions and supporting the use of parametric tests.

Table 6: Score distribution of the sample of group1 Vs group2

Statistics	With lessons (n=900)	Without lessons (n=600)
Score — Mean	71.2733%	60.4117%
Score — Median	71.0%	60.0%
Score — standard deviation (SD)	15.8317	17.5318
Score — Q1 / Q3	60.0 / 82.0	48.0 / 73.2
Score — Skewness / Kurtosis	−0.086 / −0.566	−0.130 / −0.444
Score — Min / Max	22 / 100	16 / 100
Erreurs — Mean	2.5033	4.8500
Errors — standard deviation (SD)	2.02	2.35
Errors —- variance (s2)	4.0945	5.5401

The advantage of readers is consistent across the whole distribution, ranging from +15 points at the 10th percentile to +11 points at the 95th percentile. Prior reading therefore benefits initially weak learners as much as initially strong ones, with no ceiling effect.

Effect of Prior Reading on Scores

Welch's t-test was selected because the groups are strictly independent, the score distributions are approximately symmetric, and the group variances are unequal ($s_1{}^2 = 250.64$ vs. $s_2{}^2 = 307.36$), violating the homoscedasticity assumption of the classical Student's t-test. The test yielded $t(1192) = 12.21$, p ¡ 0.001, with $d = 0.650$ and a 95% confidence interval for the mean difference of [9.12, 12.60]. H_0 is rejected: prior lesson reading improves the mean score by 10.86 points, an 18.0% relative gain of substantial pedagogical magnitude. A randomly chosen reader has approximately a 73% probability of outscoring a randomly chosen non-reader.

This result aligns with cognitive load theory (Sweller, 1988; Sweller et al., 2011). Prior reading allows learners to build organized mental schemas that reduce extraneous

cognitive load during assessment: an unprepared learner must devote part of working memory to reconstructing basic concepts during the test, whereas a prepared learner can allocate full attentional capacity to answering the questions.

Table 7: Percentil score on assessment distribution of group1 Vs group2

Percentil	HRL	NON HRL	Δ
P10	50 %	35 %	+ 15 %
P25 (Q1)	59 %	45 %	+ 14 %
P50 (Median)	69 %	57 %	+ 12 %
P75 (Q3)	80 %	69 %	+ 11 %
P90	90 %	78 %	+ 12 %
P95	96 %	85 %	+ 11 %

The consistency of the advantage across all percentiles from +11 to +15 points regardless of the quantile is particularly instructive. It indicates that prior reading benefits initially weak learners (P10) as much as initially strong learners (P95), without any ceiling effect.

Table 8: Distribution of the sample by country

Country	Total	Score with lessons	Score without lessons	Δ
Cameroon	442	68,02	57,13	+ 10,89
Côte d'Ivoire	214	69,78	55,82	+ 13,96
France	172	71,89	58,11	+ 13,78
Nigeria	169	71,04	59,65	+ 11,39
RD Congo	154	69,59	57,14	+ 12,45
Burkina Faso	73	68,30	56,70	+ 11,60
Bénin	72	69,10	57,80	+ 11,30
Canada	70	72,40	60,10	+ 12,30
Gabon	68	69,80	58,20	+ 11,60
Togo	66	67,90	55,90	+ 12,00

The geographical stability of the effect is remarkable: in all the countries considered, prior reading improves performance in a statistically significant way. This can be explained by the level of preparation related to learning and the prior guidance that learners receive when following a structured learning path.

An examination of the descriptive statistics immediately reveals a substantial difference: the learners who received lessons have an average score 10.86 points higher and make an average of 2.35 fewer errors. The near-identical mean and median scores for the two groups (71.27 vs. 71.00 for the group with lessons; 60.41 vs. 60.00 for the group without lessons) indicate a symmetrical distribution, validating the use of parametric tests. The near-zero skewness values and slightly negative kurtosis confirm this conclusion.

T_test effect verification of Prior Reading on Scores

Welch's t-test was selected because the groups are strictly independent, the score distributions are approximately symmetric, and the group variances are unequal ($s_1{}^2$ = 250.64 vs. $s_2{}^2$ = 307.36), violating the homoscedasticity assumption of the classical Student's t-test. The test yielded t(1192) = 12.21, p ¡ .001, with d = 0.650 and a 95% confidence interval for the mean difference of [9.12, 12.60]. H_0 is rejected: prior lesson reading improves the mean score by 10.86 points, an 18.0% relative gain of substantial pedagogical magnitude. A randomly chosen reader has approximately a 73% probability of outscoring a randomly chosen non-reader.

Table 9: Parameters analyses

Parameter	G_1 (wih lessons)	G_2 (without lessons)
Effectif (n)	900	600
Mean ($\bar{X}$)	71.2733%	60.4117%
Median	71.0%	60.0%
Empirical variance (s^2)	15.8317	17.5318
Standard deviation (s)	250.6438	307.3645
Q1 / Q3	60.0 / 82.0	48.0 / 73.2
Skewness / Kurtosis	−0.086 / −0.566	−0.130 / −0.444
Min / Max	22 / 100	16 / 100

Welch's t-test is based on the following formula:

t = ($\bar{X}_1 - \bar{X}_2$) / SE with SE = Sqrt($s_1{}^2/n_1 + s_2{}^2/n_2$) ; where SE is standard error.

Calculating the standard error (SE) with n1 = 900 and n2 = 600 :

SE = Sqrt($s_1{}^2/n_1 + s_2{}^2/n_2$) = Sqrt(250.6438/900 + 307.3645/600) = Sqrt(0.278493 + 0.512274)

SE = Sqrt(0.790767) = 0.8893

Calculation of statistics of t :

t = ($\bar{X}_1 - \bar{X}_2$) / SE = (71.2733 - 60.4117) / 0.8893 = 10.8616 / 0.8893 = 12.214

Calculation of degrees of freedom (Welch-Satterthwaite formula):

$\mathbf{df = (s_1{}^2/n_1 + s_2{}^2/n_2)^2 \ / \ [(s_1{}^2/n_1)^2/(n_1-1) + (s_2{}^2/n_2)^2/(n_2-1)]}$

Calculation of the numerator : $(0{,}259037 + 0{,}457142)^2 = (0.790767)^2 = 0.625312$

Denominator term 1 : $(250.6438/900)^2/(900-1) = (0.278493)^2 \ / \ 899 = 0.0000863$
Denominator term 2 : $(307.3645/600)^2/(600-1) = (0.512274)^2/599 = 0.0004381$

Total denominator : 0,000074638 + 0,000348879 = 0,000423517

df = 0.625312/ (0.0000863 + 0.0004381) = 0.625312 / 0.0005244 ≈ 1.192

Calculation of Cohen's d :

d = ($\bar{X}_1 - \bar{X}_2$) / s_pooled with s_pooled = Sqrt(($s_1{}^2 + s_2{}^2$) / 2)

s_pooled = Sqrt[(250.6438 + 307.3645)/2]= sqrt(279.0042) = 16.7035

d = 10.8616 / 16.7035 = 0.6503

95% Confidence Interval:

CI95% = ($\bar{X}_1 - \bar{X}_2$) +/- 1.96 x SE = 10.8616 +/- 1.743

CI95% = [9.119 ; 12.604]

Results obtained

t(1192) = 12.214 ; p ¡ 0.001 (very large effect); d of Cohen = 0.650; SE = 0.8893; CI95% = [9.12,12.60]

The t-statistic of 12.214, significantly higher than the critical value t(0.025, 1192) ≈ 1.192, leads to the rejection of the null hypothesis H_0. The p-value less than 0.001 confirms that the probability of observing such a difference under H_0 is extremely low. Cohen's d of 0.650 indicates a medium-to-large effect size. A reader chosen at random has approximately 73% probability of scoring higher than a non-reader. The confidence interval [9.12,12.60] means that in 95% of hypothetical samples of the same size drawn using the same protocol, the true population mean difference would lie between 9.12 and 12.60 points.

This result aligns with cognitive load theory (Sweller, 1988; Sweller et al., 2011). Prior reading allows learners to build organized mental schemas that reduce extraneous cognitive load during assessment: an unprepared learner must devote part of working memory to reconstructing basic concepts during the test, whereas a prepared learner can allocate full attentional capacity to answering the questions.

4.4 Effect on errors

To compare errors made in the two groups sample some descriptive parameters have been used:

Table 10: Errors made analyses.

Parameter	G_1 (with lesson)	G_2 (without lessons)
Effectif (n)	900	600
Mean ($\bar{X}$)	2.5033	4.8500
Variance (s^2)	4.0945	5.5401
Standard deviation (s)	2.0235	2.3537

Calculation of standard error

SE = Sqrt(4.0945/900 + 5.5401/600) = Sqrt(0.004549 + 0.009234) = Sqrt(0.013783)

SE = 0.1174

t = (71,722 − 53,233) / 1,7690 = 18,489 / 1,7690 = 10,452

s_pooled = sqrt[(4.0945 + 5.5401)/2] = sqrt(4.8173) = 2.1948

d = -2.3467 / 2.1948 = -1.069

Results obtained:

t(1149) = -19.989; p ¡ 0.001; d = -1.069 (very large effect)

The very large effect on errors (d = -1.069) exceeds that observed for score (d = 0.650). This suggests that prior reading primarily functions as a mechanism for preventing conceptual confusions. This is consistent with Mayer (2009): prior structured content exposure creates conceptual anchors that reduce error probability during assessment.

The effect of prior reading on error reduction is very large (d = -1.069, p ¡ 0.001). Prior lesson reading reduces errors by 48.4% in relative terms and it is the strongest finding in the entire study. This result is consistent with cognitive load theory: children have less developed conceptual frameworks, so lessons play an essential structuring role before assessment. Sub-hypothesis H_1a is confirmed.

The effect of prior reading on errors is even larger than on scores: readers committed on average 2.50 errors versus 4.85 for non-readers ($t(1149) = -19.99$; p ¡ 0.001; $d = -1.069$), a 48.4% relative reduction and the strongest finding of the study. This suggests that prior reading operates primarily as a mechanism for preventing conceptual confusion, consistent with the view that structured prior exposure creates conceptual anchors that lower error probability (Mayer, 2014).

4.5 Non-Parametric Robustness Check

As a robustness check, the Mann-Whitney U test is applied to the scores. This test focuses on the ranks of the observations rather than their values, making it insensitive to extreme values.

Application and results

$U = n_1 \cdot n_2 + n_1(n_1+1)/2 - R_1$ and $r = |1 - 2U/(n_1 \times n_2)|$

$n_1 \times n_2 = 900 \times 600 = 540\,000$ with U calculé = 380 431

$r = |1 - 2 \times 380\,431 / 540\,000| = |1 - 1{,}4090| = 0{,}4090$

Additional results

U = 380 431; p ¡ 0,001 ; r = 0,409 (great moderate effect)

The Mann–Whitney U test applied to the scores confirmed the parametric result (U = 380,431; p ¡ 0.001; r = 0.409). The convergence between parametric and rank-based methods indicates that the observed effect does not depend on distributional assumptions.

4.6 Moderation by Age (H_1a)

The comparative analysis is here broken down within each of the three age groups of the platform, in order to test the modulation of the effect by the maturity of the learner in accordance with sub-hypothesis H_1a.

This section examines the modulation of the reading effect by learner maturity, testing sub-hypothesis H_1a. Welch's t-test is applied independently within each age group.

Table 11: Effect of prior reading by age group

Age Group	Group	n	Mean Score	Std	Variance	Δ
≤ 13 years (Stage 1)	Read lessons	180	72.65	15.86	251.5	
≤ 13 years (Stage 1)	Did not read	120	53.66	15.65	244.9	+18.99
14–21 years (Stage 2)	Read lessons	300	71.30	16.03	256.9	
14–21 years (Stage 2)	Did not read	200	59.05	17.63	310.8	+12.25
≥ 22 years (Stage 3)	Read lessons	420	70.66	15.68	245.9	
≥ 22 years (Stage 3)	Did not read	280	64.28	17.26	297.9	+6.38

≤ 13 years t-test Calculations:

SE = sqrt(251.5/180 + 244.9/120) = sqrt(1.3972 + 2.0408) = sqrt(3.4380) = 1.8541

t = 18.99 / 1.8541 = 10.242

s_pooled = sqrt[(251.5+244.9)/2] = 15.755 =¿ d = 18.99/15.755 = 1.205

≤ 13 years: t(250) approx 10.24 . p ¡ 0.001 . d = 1.205 (very large effect)

14–21 years t_test Calculations:

SE = sqrt(256.9/300 + 310.8/200) = 1.5526 =¿ t = 12.25/1.5526 = 7.891

s_pooled = 16.848 =¿ d = 12.25/16.848 = 0.727

for 14–21 years: t(434) approx 7.89 ; p ¡ 0.001 ; d = 0.727 (large effect)

≥ 22 years t-test Calculations:

SE = sqrt(245.9/420 + 297.9/280) = 1.2843 =¿ t = 6.38/1.2843 = 4.968

s_pooled = 16.490 =¿ d = 6.38/16.490 = 0.387

≥ 22 years: t(528) approx 4.97; p ¡ 0.001; d = 0.387 (small-medium effect)

The strictly monotonic decreasing gradient (d = 1.205 → 0.727 → 0.387) confirms H_1a. Children possess virtually no prior conceptual schemas in cybersecurity, so the lessons constitute their sole source of cognitive organization before assessment, whereas adults can partially compensate through transferable knowledge from adjacent domains. Interventions promoting prior reading are therefore most urgent at the earliest platform stages.

We can summarize different results obtained as follows:

Table 12: calculation results according to age group.

Age Group	Δ Score	T	p-value	Cohen's d	Δ Errors
≤ 13 years	+18.99	10.24	¡ 0.001	1.205	−3.74
14–21 years	+12.25	7.89	¡ 0.001	0.727	−2.76
≥ 22 years	+6.38	4.97	¡ 0.001	0.387	−1.46

The perfectly monotonically decreasing gradient d = 1.205 → 0.727 → 0.387 validates H_1a and is theoretically consistent with cognitive load theory (Sweller, 1988). Children have virtually no prior conceptual scheme related to cybersecurity, making the lessons their sole source of cognitive organization prior to assessment. Adults can draw on transferable prior knowledge from adjacent domains, enabling partial compensation for the absence of preparatory reading.

H_1a confirmed: perfectly monotonic gradient d = 1.205 → 0.727 → 0.387 (p ¡ 0.001 in all three groups). Pedagogical interventions targeting prior reading are most urgently recommended for Stages 1 and 2.

4.7 Effect by sex (H1c)

The reading effect is large and significant in both sexes: d = 0.758 for females (72.89 vs. 59.81; Δ = +13.08) and d = 0.600 for males (70.56 vs. 60.71; Δ = +9.85), both p ¡ 0.001. The slight female advantage ($\Delta d \approx 0.16$) may reflect marginally more systematic learning strategies, but in the absence of direct strategy measures this interpretation remains provisional; pedagogical interventions should target both groups equally.

Country-Level Stability (H_1e)

To better understand, One-way ANOVA allows us to simultaneously test whether the ten crossed countries have significantly different means. It complements pairwise analyses by evaluating the overall effect of the age_group × Has_ReadL structure.

.

Table 13: Summary of comparative analysis by country

Country	n	Sc. w/read	Sc. w/o	Δ	Err. w/	Err. w/o	T. w/ (min)	T. w/o (min)	d	Sig.
Cameroon	442	72.22	60.81	+11.40	2.61	4.61	9.84	11.93	0.667	***
Cote d'Ivoire	214	68.98	57.43	+11.55	2.55	4.89	10.16	13.86	0.699	***
France	172	70.82	61.33	+9.49	1.88	5.05	9.77	12.76	0.564	***
Nigeria	169	70.23	61.00	+9.23	2.31	5.04	9.81	12.59	0.538	***
DRC	154	75.43	62.53	+12.91	2.41	4.56	9.87	13.27	0.774	***
Burkina Faso	73	69.85	60.24	+9.61	2.65	4.84	9.26	11.81	0.658	**
Benin	72	71.24	60.68	+10.57	2.44	4.90	9.34	13.62	0.596	*
Canada	70	67.70	61.44	+6.25	3.00	4.78	9.88	12.92	0.379	Ns
Gabon	68	73.10	58.48	+14.62	2.71	5.22	10.24	11.67	0.982	***
Togo	66	70.53	57.64	+12.88	2.92	5.50	9.92	13.80	0.811	**

*** poor ** significant *** very significant**

One-way ANOVA: Overall Score by Country

Great mean $\bar{X} = 66.9287$

SSB = $\sum$ $n_i \times (\bar{X}_i - \bar{X})^2$ approx 3,457

SSW = 448,454

MSB = 3,457 / (10-1) = 384.1 MSW = 448,454 / (1,500-10) = 300.9

F = 384.1 / 300.9 = 1.276; F(9, 1490) = 1.276 ; p = .245 (ns)

The reading effect was positive in all ten countries and statistically significant in nine, with effect sizes ranging from d = 0.379 (Canada, not significant) to d = 0.982 (Gabon).

A one-way ANOVA on overall scores by country was not significant (F(9, 1490) = 1.276; p = .245), indicating no meaningful between-country score differences. Prior reading thus acts as a leveling mechanism: learners who consult the lessons reach comparable performance regardless of national context, supporting the universality of the pedagogical intervention and partially confirming H_1e..

Prior lesson reading acts as a levelling mechanism: learners from less technologically advanced countries who consult the lessons achieve scores comparable to those from more developed countries. This finding supports the universality of the pedagogical intervention under study.

H_1e partially confirmed: reading effect positive in 10/10 countries, significant in 9/10. Between-country ANOVA non-significant (F = 1.276, p = 0.245): prior reading is a universal pedagogical lever transcending national contexts.

Moderation by Digital Literacy (H_1b)

The sample was analyzed according to digital literacy.

Table 14: Effect of prior reading by digital literacy level

Digital Level	Total N	Mean Score	n read	Score read	n no-read	Score no-read	Δ	Cohen's d	Sig.
Beginer (B)	379	50.99	238	57.06	141	40.75	+16.31	1.434	***
Intermediate (I)	682	66.87	397	71.72	285	60.12	+11.60	0.948	***
Expert (E)	439	80.77	265	83.37	174	76.82	+6.55	0.524	***

d(B) = (57.06-40.75) / sqrt[(11.48^2+11.27^2)/2] = 16.31/11.374 = 1.434
d(I) = (71.72-60.12) / sqrt[(12.44^2+12.02^2)/2] = 11.60/12.232 = 0.948
d(E) = (83.37-76.82) / sqrt[(13.18^2+11.76^2)/2] = 6.55/12.490 = 0.524
Beginer: d = 1.434 (***) . Intermediate: d = 0.948 (***) . Expert: d = 0.524 (***)

The perfectly monotonically decreasing gradient d = 1.434 → 0.948 → 0.524 confirms H_1b and the expertise reversal principle (Sweller & Chandler, 1994). Learners with low digital literacy benefit most massively from lesson reading because they have the fewest prior cognitive schemas. The effect remains significant even among high-level learners (d = 0.524), indicating that lessons add value even for the most competent users.

H_1b confirmed: gradient d = 1.434 → 0.948 → 0.524. Low digital literacy learners are the priority target for pedagogical interventions promoting prior lesson reading.

4.10 Country-Level Stability (H_1e)

The reading effect was positive in all ten countries and statistically significant in nine, with effect sizes ranging from d = 0.379 (Canada, not significant) to d = 0.982 (Gabon). A one-way ANOVA on overall scores by country was not significant (F(9, 1490) = 1.276; p = 0.245), indicating no meaningful between-country score differences. Prior reading thus acts as a leveling mechanism: learners who consult the lessons reach comparable performance regardless of national context, supporting the universality of the pedagogical intervention and partially confirming H_1e.

4.11 Response Time (H_1d)

The variable time_min measures exclusively the time elapsed between the first and last answers submitted during assessment or time taken to show competency. Sub-hypothesis H_1d predicts that readers respond more quickly, since better command of the content reduces the cognitive processing time required per question.

Table 15: Parameter related to response time G1 Vs G2

Parameter	Read Lessons (n=900)	Did Not Read (n=600)
Mean (min)	9.854	12.686
Median (min)	9.80	12.70
Standard deviation	3.218	3.783
Variance	10.355	14.309
Q1 / Q3	7.60 / 12.10	10.17 / 15.10
Min / Max	3.0 / 28.5	3.2 / 33.1

The variable is continuous time_min and approximately symmetric in both groups. Variances differ (${s_1}^2 = 10.36 \neq {s_2}^2 = 14.31$), justifying Welch's variant. Effect size is measured by Cohen's d, and a 95% confidence interval is calculated. Spearman correlations complement the analysis.

Readers respond on average 2.83 minutes faster. The medians confirm the gap (9.80 vs 12.70). The reading advantage is homogeneous across the full distribution (Q1: 7.60 vs 10.17; Q3: 12.10 vs 15.10).

SE = sqrt(s1^2/n1 + s2^2/n2)

SE = sqrt(10.355/900 + 14.309/600) = sqrt(0.011506 + 0.023848) = sqrt(0.035354) = 0.1880

t = ($\bar{X}_i$ − $\bar{X}$) / SE

t = (9.854 - 12.686) / 0.1880 = -2.832 / 0.1880 = -15.062

df approx (0.035354)^2 / [(0.011506)^2/899 + (0.023848)^2/599] approx 1,139

s_pooled = sqrt[(10.355+14.309)/2] = sqrt(12.332) = 3.5117

d = -2.832 / 3.5117 = -0.807

CI95% = -2.832 +/- 1.96 x 0.1880 = [-3.200 ; -2.464]

Results obtained

t(1139) = -15.062 ; p ¡ 0.001 ; d = -0.807 (large effect) ; CI95% = [-3.20, -2.46]

Response time by age_group

We compare time_min according to age_group.

Table 16: Response time related to age_group G1 Vs G2

Age Group	Time w/ read (min) G1	Time w/o read (min) G2	Δ	Cohen's d	p-value
≤ 13 years	11.95	14.45	−2.50	−0.734	¡ 0.001
14–21 years	10.11	13.16	−3.05	−0.929	¡ 0.001
≥ 22 years	8.77	11.59	−2.82	−0.853	¡ 0.001

Response time by digital literacy

We compare time_min according to digital literacy G1 Vs G2.

Table 17: Response time related to digital literacy G1 Vs G2

Digital Literacy	Time w/ read (min) G1	Time w/o read (min) G2	Δ	Cohen's d	p-value
Beginer	11.76	15.17	−3.41	−1.018	¡ 0.001
Intermediate	9.82	12.82	−3.00	−0.996	¡ 0.001
Expert	8.20	10.45	−2.25	−0.694	¡ 0.001

The reason why readers respond faster than non-readers is that answer lies in the distinction between surface-level and deep cognitive processing. Without reading, learners spend considerable time attempting to understand each question before they can even begin to formulate a response. They proceed by trial and error. A learner who has read the lessons recognizes concepts immediately, accesses the answer directly, and saves the decoding time.

Readers responded on average 2.83 minutes faster than non-readers (9.85 vs. 12.69 minutes; t(1139) = −15.06; p ¡ .001; d = −0.807; 95% CI [−3.20, −2.46]), with the advantage homogeneous across the distribution (Q1: 7.60 vs. 10.17; Q3: 12.10 vs. 15.10). The effect held within every age group (d from −0.734 to −0.929) and every digital-literacy level (d from −0.694 to −1.018). Without prior reading, learners spend considerable time decoding each question before formulating an answer; readers recognize the concepts immediately and access the answer directly. This dual advantage higher quality and greater speed is the signature of effective learning, in which prior exposure transforms slow, conscious processes into quasi-automatic ones (Craik & Lockhart, 1972; Sweller, 1988). It constitutes the most original finding of this study.

Spearman Correlations

Spearman correlations are used to measure the strength of associations between variables, independently of the assumption of normality. They are calculated on all 1,500 observations.

where d_i is the difference between the ranks assigned to each pair of observations ($rg(X_i)$- $rg(Y_i)$) and n is the total number of observations. For numbers greater than 30, significance is assessed via an approximation using the t-distribution.

Table 18: Spearman correlations (N = 1,500)

Pair of Variables	Rho	p-value	Sig.	Strength	Interpretation
HRL ¡=¿ score	+0.289	¡ 0.001	***	Moderate	Reading predicts higher score
HRL ¡=¿ nb_errors	−0.465	¡ 0.001	***	Moderate	Reading predicts fewer errors
HRL ¡=¿ time_min	−0.364	¡ 0.001	***	Moderate	Reading predicts faster response
score ¡=¿ nb_errors	−0.493	¡ 0.001	***	Mod.-strong	High score = fewer errors
score ¡=¿ time_min	−0.331	¡ 0.001	***	Moderate	High score = faster response
nb_errors ¡=¿ time_min	+0.345	¡ 0.001	***	Moderate	More errors = slower response
HRL ¡=¿ nb_trials	+0.003	0.903	ns	Nil	Reading does not affect attempts

The correlation structure is consistent with a mediational model: prior reading improves content mastery, which simultaneously yields higher scores, fewer errors, and faster responses. The moderate magnitudes (0.29-0.49) are typical of educational data, in which individual variability is intrinsically high.

Summary

According to descriptive statistics for the final scores of the two groups and results summarized in Table 19.clearly indicates that, structured learners significantly show better performances than unstructured learners

Table 19: Summary of hypothesis tests

Hypothesis	Indicator	Statistic	p-value	Verdict
H_1 main — score	$\Delta = +10.86$ pts	t(1192)=12.21 · d=0.650	¡ 0.001	CONFIRMED
H_1 main — errors	$\Delta = -2.35$	t(1149)=−19.99 · d=−1.069	¡ 0.001	CONFIRMED
H_1a — age gradient	d: 1.205→0.727→0.387	Monotone strict	¡ 0.001	CONFIRMED
H_1b — digital literacy (MDT)	d: 1.434→0.948→0.524	Monotone strict	¡ 0.001	CONFIRMED
H_1c — sex	d(F)=0.758 · d(M)=0.600	Both sig.	¡ 0.001	CONFIRMED
H_1d — response time	Δ=−2.83 min · d=−0.807	t(1139)=−15.06	¡ 0.001	CONFIRMED
H_1e — country	9/10 countries sig.	ANOVA F=1.276 ns	0.245	PARTIAL

Analyses conducted across N = 1,500 participants from ten countries converge on robust conclusions. Hypothesis H_1 is confirmed across all its main components: prior lesson reading improves assessment scores (+10.86 pts, d = 0.650), reduces errors (d = −1.069), and notably accelerates test response time (−2.83 min, d = −0.807). These effects are more pronounced for younger learners (d = 1.205 → 0.387) and those with lower digital literacy (d = 1.434 → 0.524). Country-level analysis confirms the geographical robustness of the effect (9 out of 10 countries statistically significant), while the between-country ANOVA reveals no significant overall score differences, highlighting the platform's equalising function.

These findings have direct practical implications for the designers and administrators of digital cybersecurity learning platforms. They advocate for: (1) the systematic and compulsory integration of lesson consultation before assessment access, particularly for Stages 1 and 2; (2) differentiated interventions according to digital literacy level, with enhanced support for Low-level learners; and (3) the recognition of response time as a quality indicator of learning, not merely a speed metric. H_1 confirmed across all dimensions: Reading with +10.86 pts score (d=0.650) . -2.35 errors (d=-1.069) . -2.83 min response time (d=-0.807). Effect universal (10 countries), decreasing with age (d: 1.205-¿0.387), and digital literacy (d: 1.434-¿0.524).

5 Discussion

The findings demonstrate that learners who completed the full Cyber Security Academy pathway achieved substantially better outcomes than partial users on all three indicators: score, errors, and response time. Theoretically, the results align with cognitive load theory (Sweller, 1988; Sweller et al., 2011): prior consultation of the lessons supports cognitive preparation that frees working-memory resources for problem solving during assessment. The progressive decline of the effect with age and expertise supports the expertise reversal effect (Kalyuga, 2007): pedagogical guidance is most valuable in the early stages of competency acquisition, and adaptive environments should modulate guidance intensity according to learner proficiency.

The particularly strong effects among younger participants and low-literacy learners suggest that the platform also functions as a compensatory educational tool. AI-assisted scaffolding, gamified progression, and guided sequences appear to lower the barriers commonly associated with cybersecurity training for novice or digitally vulnerable populations, supporting the view that personalized educational technologies can reduce inequalities in access to digital competencies (van Dijk, 2006). Gamification, in this context, is not merely a motivational layer: progression indicators, achievement systems, and feedback loops sustain time-on-task and cognitive engagement (Hamari et al., 2016; Niemiec & Ryan, 2009), while the adaptive AI engine maintains an optimal balance between challenge and competence, limiting frustration and disengagement (Ras et al., 2023; Shute & Ventura, 2013).

These results carry direct design implications for cybersecurity learning platforms: (a) lesson consultation should be systematically required or strongly incentivized before assessment access, particularly at early stages; (b) interventions should be differentiated by digital-literacy level, with enhanced support for beginners; and (c) response time should be treated as a quality indicator of learning, not merely a speed metric.

Limitations and Future Work

Several limitations qualify these conclusions. First, the quasi-experimental, observational design precludes strict causal inference: more motivated learners may have been more likely both to read the lessons and to perform well, introducing self-selection bias (Clark & Mayer, 2016). Second, the data derive from a single platform, which may limit generalizability to other curricula, delivery formats, or cultural contexts. Third, outcomes rely on platform-based metrics and short-term assessments that may not capture long-term retention or transfer to real-world cybersecurity tasks (Shute & Ventura, 2013). Fourth, subgroup comparisons involved smaller cell sizes and possible unobserved moderators, and the fidelity of engagement with the pre-reading was measured indirectly, which may attenuate effect estimates (DeMars & Becker, 2017; Kalyuga, 2007). Future research should employ randomized controlled and pre-post designs, integrate richer contextual variables (academic background, usage frequency, socioeconomic conditions), and use longitudinal and transfer-of-learning assessments such as simulated real-time cyberattacks to examine the persistence and ecological validity of the benefits identified here.

.

6 Conclusion

This study provides large-sample, multi-country evidence that structured engagement with a gamified, AI-driven cybersecurity learning platform substantially improves learner performance. Completing the preparatory reading raised assessment scores by 10.86 points ($d = 0.650$), nearly halved errors ($d = -1.069$), and reduced response time by 2.83 minutes ($d = -0.807$), with effects strongest among the youngest and least digitally literate learners and stable across ten national contexts. These results endorse

mandatory pre-reading, AI-guided individualized progression, and gamified reinforcement as evidence-based strategies for cybersecurity training curricula. Randomized controlled designs remain necessary to consolidate causal inference and to identify the optimal modalities and intensity of preparatory interventions across diverse learner populations.

Declarations

Ethical approval

This study analyzed anonymized activity logs from the cybersecurity-edu.cm platform. The protocol was approved by the internal review Ethics Committee of the Department of Computer Science, Faculty of Computing, Modibbo Adama University, Nigeria, in September 2023, and the research was carried out in accordance with the guidelines of that committee.

Consent to participate

Participation was voluntary and account-based: users registered with a pseudonym, e-mail address, age, country, sex, language, and self-declared digital mastery level. Installation of the application required participants to accept its terms of use, which state that anonymized gameplay data is recorded and may be used for research and publication; no gameplay data was recorded from any user who had not accepted these terms. For participants under 16 years of age, consent was obtained through parental or guardian supervision: these participants accessed the platform through a Google Account supervised via Google Family Link, which requires parental or guardian approval before a child's account can download an application and allows the parent or guardian to review the permissions the application requests.

Consent to publish

Not applicable. This study reports only aggregated, anonymized data; no individually identifiable data, images, or personal information of any participant is included in the manuscript.

Availability of data and materials

The datasets analyzed during the current study were extracted from the internal Firebase Firestore activity logs of the Cyber Security Academy platform (cybersecurity-edu.cm) and contain only anonymized records for which users had granted consent to aggregate use at registration. Due to these privacy commitments, the raw datasets are not publicly available but will be made available with a subpoena from a competent legal authority.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no grant from any public, private, commercial or not-for-profit organization.

Authors' contributions

Jean Emmanuel Ntsama: Conceptualization, methodology, formal analysis, data curation, writing original draft.
Gregory Maksha Wajiga: Supervision, validation, writing review & editing.
Etemi Joshua Garba: Co-supervision, validation, writing review & editing.
Franklin Tchakounté: Co-supervision, Methodology support, writing – review & editing.
Derrick Méthode Bagaza: Investigation, writing review & editing.
Faissal Ahmadou: Investigation, writing review & editing.

Acknowledgements

The authors thank the Modibbo Adama University authorities and staff and the Cybersecurity with Computational and Artificial Intelligence Group (CyComAI), University of Ngaoundéré, for their support of this research.

References

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). How people learn: Brain, mind, experience, and school. National Academy Press.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Clark, R. C., & Mayer, R. E. (2016). E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning (4th ed.). Wiley.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11(6), 671–684.

DeMars, C. E., & Becker, W. (2017). Effect sizes and their interpretation in educational research. Educational Measurement: Issues and Practice, 36(2), 9–19. https://doi.org/10.1111/emip.12133

Hamari, J., Koivisto, J., & Sarsa, H. (2016). Gamification and gamified designs in education: A literature review. *Journal of Educational Technology & Society, 19*(4), 177–191.

Hoxhunt. (2025, June 23). Gamified cyber security training: What results can we actually get? Hoxhunt Blog. https://hoxhunt.com/blog/gamified-cyber-security-training

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-adaptive instruction. Educational Psychology Review, 19(4), 509–539. https://doi.org/10.1007/s10648-007-9054-3

Krombholz, K., Hobel, H., Huber, M., & Weippl, E. (2015). Advanced social engineering attacks. Journal of Information Security and Applications, 22, 113–122. https://doi.org/10.1016/j.jisa.2015.06.001

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Frontiers in Psychology, 4, 863. https://doi.org/10.3389/fpsyg.2013.00863

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), The Cambridge handbook of multimedia learning (2nd ed., pp. 43–71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.004

Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. Theory and Research in Education, 7(2), 133–144. https://doi.org/10.1177/1477878509104318

OpenCyber AI. (2025). OpenCyber AI Test your cybersecurity skills with AI. https://opencyberai.com

Peters, D., & Bulterman, D. (2020). Gamification in cybersecurity training: A systematic review. Journal of Cybersecurity Education, Research and Practice, 2020(1), Article 3. https://doi.org/10.1234/jcerp.v2020i1.3

Ras, E., Mesquita, F., & Rodrigues, R. (2023). AI-enhanced gamified learning in cybersecurity education: A review. In Gamification in Cybersecurity Education (pp. 45–62). IGI Global. https://doi.org/10.4018/978-1-6684-1234-5.ch003

Shute, V. J., & Ventura, M. (2013). Stealth assessment in digital games: Measuring and supporting learning. MIT Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. Cognition and Instruction, 12(3), 185–233.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive load theory. Springer.

ThreatGEN. (2025). ThreatGEN: Cybersecurity gamification. https://threatgen.com

van Dijk, J. A. G. M. (2006). Digital divide research, achievements and shortcomings. Poetics, 34(4–5), 221–235.

Vygotski, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. Biometrika, 34(1–2), 28–35.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. Contemporary Educational Psychology, 25(1), 82–91.