Author Proof

Check for updates

Efficiency and Equity in Screening for Participants in Representative Cybersecurity Studies

Po-Yu Chen[1], Jinwoo Oh[1], Hsiang-Wen Hsing[1], Nathan Lau[1(✉)], Brandon Beltz[2], Peggy Wu[3], and Quanyan Zhu[4]

[1] Virginia Tech, Blacksburg, VA 24061, USA
nkclau@vt.edu
[2] The Bulls Run Group, Bethesda, MD 20817, USA
[3] Raytheon Technologies, Arlington, VA 22209, USA
[4] New York University, New York, NY 10012, USA

Abstract. Human factors research in cybersecurity confronts the practical challenge of how to recruit and screen participants with the skills necessary to perform demanding, ecologically valid tasks. Credible results from hacking studies rely on participants possessing sufficient technical skills to maneuver inside a representative network in a cyber range. Screening via self-reports can be unreliable, while rigorous skills testing is resource intensive. This paper presents empirical research investigating the effectiveness of cybersecurity certifications and cyber range skills tests for screening domain-specific competence. The Guarding Against Malicious Biased Threats (GAMBiT) project recruited participants with a two-stage screening process. The project screened 269 candidates for a final sample of 61 qualified participants. We compared the capability between the Offensive Security Certified Professional (OSCP) certification and a custom Capture-the-Flag (CTF) assessment ("Cedar Bunny") at predicting actual performance in a two-day cyber range experiment. Results indicate that OSCP holders scored significantly higher on the skills test ($p < 0.001$) and progressed further in the hacking experiment ($p = 0.002$) than non-holders. Skills test scores correlated moderately and positively with hacking progress ($\rho = 0.344$). Furthermore, hierarchical regression analysis demonstrated that the skills test provided incremental predictive validity ($\Delta R^2 = 5.2\%$) beyond certification alone. These findings suggest that while certifications are efficient proxies and skills tests can be effective supplements for identifying viable, albeit non-certified, participants. We propose a mixed screening approach to balance recruitment efficiency and participant diversity.

AQ1

Keywords: Cybersecurity · Human Factors · Participant Screening · OSCP · Skills Assessment

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
A. Moallem (Ed.): HCII 2026, LNCS 16738, pp. 1–14, 2026.
https://doi.org/10.1007/978-3-032-29723-5_24

Author Proof

1 Introduction

The validity of human factors research depends heavily on the quality of its participants. Unlike studies on general phishing susceptibility or password hygiene that can recruit from the general population, research on offensive cyber operations requires participants with specialized technical skills. To produce any meaningful results, "red team" or hacking studies must ensure that participants have sufficient command-line fluency to maneuver inside target networks, scan for vulnerabilities, and execute exploits [3]. Without this baseline competence, experimental data may reflect a lack of tool fluency rather than the cognitive behaviors or topics of interest that the study intends to assess.

This requirement creates a tension between efficiency and equity in recruitment. On one hand, researchers face an efficiency imperative. Researchers need to identify qualified participants quickly and minimize the administrative burden of screening. Subjective reports on cybersecurity knowledge and experience are efficient but sometimes unreliable for screening [1]. Participants may overstate their abilities due to the Dunning-Kruger effect [11, 13] or social desirability bias [5, 17], leading to a pool of "false eligibles" who ultimately fail to complete experimental tasks. Further, some criteria may be too broad for identifying individuals with specific technical capabilities of interests for the study. Giboney et al. (2016) found a pitfall of past studies relying on self-reported data is the inability to categorize knowledge level of the participants appropriately [8].

An alternative is relying on specific credentials or certificates for qualifying participants that should in effect provide guarantees of basic hacking skills. Recruiting individuals with certificates can be difficult and expensive. The cybersecurity workforce gap currently stands at nearly 4.8 million professionals [9], a small and competitive pool of talent. Further, screening by formal credentials (such as university degrees or expensive commercial certifications) risks excluding a significant portion of the hacker community—specifically, skilled self-taught practitioners. Hobbyist hackers might possess sophisticated operational skills honed through competitions rather than formal classrooms [19], raising an equity issue. Excluding this demographic also threatens external validity, as the resulting sample may reflect corporate security compliance rather than the adversarial mindset some studies seek to investigate [15].

On the other hand, the recruitment process can employ skills testing in cyber environments relevant for the research of interests [7]. This option may alleviate issues concerning recruitment pool size equity but is resource intensive in at least two ways. First, cybersecurity exercises are conducted in hours or days, but the preparation often cost months, making them costly and time-consuming [27]. According to Soetan [22], the manual setup time for the Capture-the-Flag (CTF) platform with challenges averages around 2 h and design of CTF challenges always faces a tradeoff between difficulty and engagement. Second, researchers need to invest time for testing participants rather than collecting useful research data.

This study, conducted under the Guarding Against Malicious Biased Threats (GAMBiT) project [2, 3], evaluates the predictive validity of cybersecurity certifications versus skills test for screening domain-aligned competence. Specifically, we analyzed screening data from 269 candidates and the experimental performance of 61 qualified participants with respect to their status on holding the Offensive Security Certified Professional (OSCP) certification and scores of the SimSpace's proprietary 'Cedar Bunny' skills

Author Proof

test. This analysis seeks to determine whether standard certifications or custom skills tests would more accurately predict hacking proficiency in research and operational environments.

2 Related Works

2.1 The Challenge of Workforce Assessment in Cybersecurity

The assessment of cybersecurity competence faces a fundamental problem: subjective ratings of competence and experience are often unreliable. Hackers may self-report as more knowledgeable to gain incentives in research studies [8]. In an observational study of 69 information technology students, Moneva et al. (2024) found no statistically significant association between self-reported hacking experience and actual hacking efficiency [15]. Although formal degrees in computer science or cybersecurity and professional certifications can be the starting point of assessment, they are criticized as point-in-time knowledge tests that fail to measure skill levels and their developments [1]. Formal education can create a validity gap where candidates may possess the vocabulary of cybersecurity without the requisite operational skills.

2.2 The Validity of Certifications: Knowledge vs. Performance

Recruiters [10, 12, 18] and researchers [14, 19] primarily rely on professional certifications for candidate screening. A recent survey of IT executives reveals that 67% of companies now mandate cybersecurity certifications as a prerequisite for employment [6]. The literature increasingly views cybersecurity education as a "three-legged stool" comprising of curriculum, hands-on skills, and certifications [26].

Certifications are separated into knowledge-based or performance-based. Knowledge-based certifications focus on policy and terminology, which may not translate to operational security capability [4]. Performance-based certifications, particularly the Offensive Security Certified Professional (OSCP), have emerged as industry benchmarks for technical skill. The OSCP requires a 23-h and 45-min practical exam where candidates must compromise target machines then followed by another 24 h to write a comprehensive report of the penetration test [16]. OSCP is internationally-recognized as the "Holy Grail" of offensive security certifications for qualifying proficiency in the hacking techniques deployed by professional penetration testers [20]. It is also the most frequently required certification for penetration testing jobs across 935 cybersecurity positions [18].

2.3 Gamification and CTFs as Psychometric Tools

CTF competitions have expanded beyond hobbyist events to function as formal recruitment tools for researchers. Researchers utilize these environments to observe technical capabilities and problem-solving behaviors [24], providing dynamic performance data that resumes or self-reports cannot capture. While past competition formats relying on isolated problems may struggle with ecological validity [23, 25], contemporary

Author Proof

CTF competitions are hosted with increasingly representative network operational environments. Soetan (2025) demonstrated that reusable, standardized CTF platforms can effectively assess candidates for specific technical job positions [22].

2.4 The Efficiency-Equity Tradeoff

Screening participants with certificates for advanced competencies, such as OSCP, or representative skills tests in a cyber range can ensure quality of participant pools. However, both present their own challenges. Advanced-level certifications are efficient indication of qualified candidates, but there is a serious risk of a limited participant pool and exclusion of "hidden talent" – self-taught hackers who lack the financial reasons or means to pursue expensive credentials [15]. Representative skills tests, such as CTF in a complex cyber range, may alleviate the risk of small participant pool and inequitable recruitment, but there is often a major cost in running such tests. There is an inherent efficiency-equity tradeoff between the two screening criteria. This reflects the dilemma in human subject research on how to secure a representative sample of skilled hackers not only efficiently but also equitably.

3 Methodology

To shed light on this efficiency-equity trade-off in recruitment, this study explores how well a cybersecurity certificate and a relatively demanding cybersecurity skills test in a cyber range predicted performance scores of a major hacking study.

3.1 Study Context and Design

The empirical research described in this paper was conducted as part of the GAMBiT project, which utilized a high-fidelity cyber range to study cognitive biases in cyber operations. The experiment required the participant to act as a "red team" operator, infiltrating a simulated corporate network to locate and exfiltrate sensitive data while navigating cognitive triggers embedded in the environment [2, 3]. The study employed a two-step screening process designed to filter a broad applicant pool down to a qualified cohort capable of completing these tasks without significant researcher intervention (Fig. 1).

3.2 Phase 1: Recruitment and Self-Report Screening

Recruitment relied on university flyers, professional networks, and social media channels targeting cybersecurity interest groups. A total of 269 individuals were screened. The initial screening questionnaire collected self-reported data on demographics, educational background, employment history, and technical skills. Interested participants must meet our initial criteria on their responses to proceed to the next screening phase (Table 1).

In this questionnaire, 149 participants disclosed their status for specific certifications: 119 participants responded to have at least one type of certification, 41 of them being Offensive Security Certified Professional (OSCP). The OSCP certification should be

Author Proof

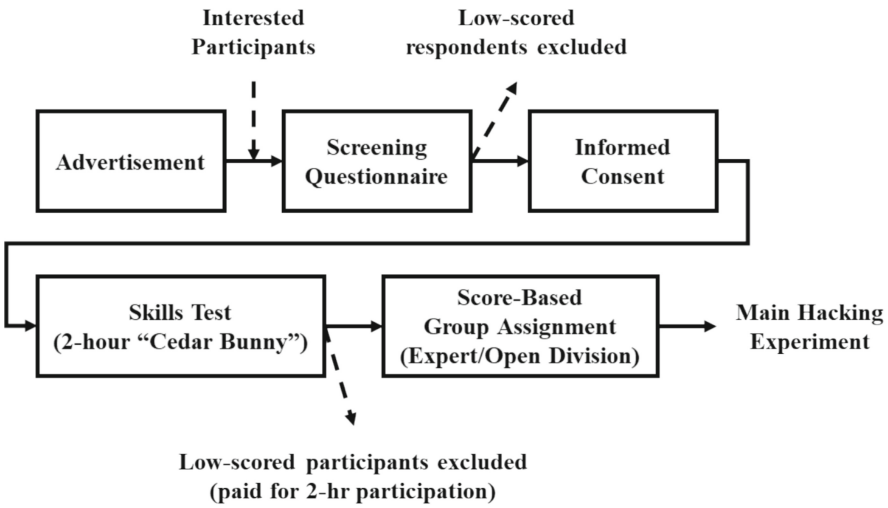


Fig. 1. Overview of recruitment process.

the most sensitive to the hacking performance of the participants because the OSCP certificate and the GAMBiT experiment focused on using the Kali Linux platform and manual exploitation techniques.

3.3 Phase 2: the "Cedar Bunny" Skills Test

From the initial pool, 149 individuals met the basic eligibility criteria and progressed to the second screening phase. This phase involved a proprietary, two-hour cybersecurity skills test titled "Cedar Bunny", hosted on the SimSpace cyber range [21]. The Cedar Bunny test was utilized to characterize participants' technical competencies and validate self-reported skill levels. This assessment combined multiple-choice questions and tasks within a standard Kali Linux environment, covering:

1. Network Enumeration: Enumerating remote services and performing standard service interrogation.
2. Vulnerability Scanning: Selecting tools to identify and leverage vulnerabilities within the target environment.
3. Exploitation: Wielding public exploits using Metasploit and searchsploit, and constructing a simple buffer overflow against a SUID-root binary.
4. Privilege Escalation: Executing maneuvers to achieve local elevations and privilege escalation.

Cedar Bunny performance scores ranged between 0 and 1200 points. To ensure that participants could perform basic hacking tasks independently for the two-day experiment, a cutoff score of 420 was established. This threshold was empirically determined during pilot testing as the minimum level of fluency on the Linux command line and execution of exploits without assistance. Of the 149 participants who attempted the test, 100 achieved the passing score, representing a pass rate of approximately 67%.

Author Proof

Table 1. Inclusion criteria for the screening questionnaire.

Eligibility	**> 18 years old** **> 1 years of hacking experience** in either hobbies or professional **> 10 h/week in hacking** in recent 1 year
Basic skills in environment	**Expressed familiarity to Kali Linux > = 3** (Not at all = 1, Extremely Familiar = 5)
Basic proficiency in core skills	**Self-rated skill level > = 2** for all skills (Don't know = 0, Novice = 1, Expert = 5) - Information Gathering - Locating Public Exploits - Tunneling and Pivoting - Using Command and Control Framework
Basic proficiency in other skills	**Self-rated skill level > = 1** for all skills (Don't know = 0, Novice = 1, Expert = 5) - Active Directory Enumeration & Attacks - Buffer Overflows - Client-Side Attacks - Enumerating Remote Services - Password Attacks & Cracking - Port Scanning & Redirection - Privilege Escalation - Scripting - SQL Injection Attacks - Target Reconnaissance & Vulnerability Scanning - Web Application Attacks
Willingness to participate	Respond '**Yes**' to the willingness to participate in two-day experiment (~8 h a day)

3.4 The Main Hacking Experiment

Not all 100 qualified participants were able to complete the experiment due to their different reasons (e.g., scheduling). At the end, 61 participants completed the full GAMBiT experiment. The experiment was a two-day "red team" operation conducted in a high-fidelity virtual and representative corporate network hosted on the SimSpace cyber range. The topology consists of around 40 virtual devices configured into subnetworks with routers, switches, including a domain controller, a file server, and several user workstations running various operating systems. The network was populated with "grey" background traffic and simulated user activity to enhance realism [2, 3].

Participants acted as external attackers using a Kali Linux jump box. Their objective was to move laterally through the network, escalate privileges, and exfiltrate "flag" files representing intellectual property. Unlike the Cedar Bunny skills test, which used a point system, this experiment assessed participants on their hacking progress on an ordinal scale from 1 to 12. This "hacking progress" score corresponded to specific target machines reached on an ideal linear attack path (i.e. score of 1 at the entry point and

Author Proof

score of 12 for the last/deepest target). This ordinal metric provided a granular view of how deep each participant penetrated the network.

3.5 Demographics

The demographics of the 61 participants who completed the main experiment is in Table 2. The age ranged between 19 and 43 years (M = 26.8, SD = 5.4). The participants were predominantly male (59 males, 96.7%; 2 females, 3.3%). Most participants reported their country of origin to be India (42.6%) or the United States (21.3%), with the remaining participants (36.1%) coming from Lebanon, Malaysia, Indonesia, Singapore, Israel, Egypt, Hong Kong, Syria, Kazakhstan, Brazil, Portugal, Taiwan, Saudi Arabia, Vietnam, and Pakistan. Approximately half reported their native language to be English (47.5%). In terms of employment, 57.4% were employed full-time, 18.0% employed part-time, 23.0% students, and 1.6% unemployed.

Many participants reported that hacking was a professional activity with 45.9% reported hacking as their primary source of income, 27.9% as a secondary source, and 23.0% reported no hacking-related income. Participants had substantial experience in professional roles requiring hacking skills, with 14.8% less than 1 year, 49.2% between 1 and 3 years, 18.0% between 4 and 6 years, and 6.6% between 7 and 10 years; 11.5% without any professional roles requiring hacking skills. For participants who were hobbyist hackers, 47.5% had engaged in hacking between 1 and 3 years, 32.8% between 4 and 6 years, 14.8% between 7. And 10 years, and 4.9% with more than 10 years. Most participants (80.3%) reported holding at least one cybersecurity certificate. Nearly all had prior experience with hackathons or CTF competitions throughout their lifetimes, with 42.6% reporting participation between 1 and 5 events, 23.0% between 6 and 10, 8.2% between 11 and 15, and 14.8% in 16 or more, while 11.5% reported no competition experience.

3.6 Hypotheses

Based on the research question, we hypothesized that:

- H1: OSCP holders would score higher on Cedar Bunny than others
- H2: OSCP holders would score higher in the two-day experiment than others
- H3: Cedar Bunny and hacking experiment progress would correlate positively
- H4: The Cedar Bunny skills test can supplement OSCP certification in predicting hacking progress.

4 Results

To investigate the relative effectiveness of the OSCP certificate and the Cedar Bunny skills test, we utilized non-parametric statistical tests due to violation of normality for the Cedar Bunny scores and the ordinal scale of hacking progress.

Author Proof

Table 2. Participant demographics and hacking experience (N = 61).

Variable	% or M (SD)
Age (years)	M = 26.8 (SD = 5.4), range 19–43
Gender	Male (n = 59; 96.7%), Female (n = 2; 3.3%)
Country of origin	India (n = 26; 42.6%), USA (n = 13; 21.3%), other countries (n = 22; 36.1%)
Native English speaker	Yes (n = 29; 47.5%), No (n = 32; 52.5%)
Employment status	Full time (n = 35; 57.4%), Part time (n = 11; 18.0%), Students (n = 14; 23.0%), Unemployed (n = 1; 1.6%)
Hacking as the source of income	Primary (n = 28; 45.9%), Secondary (n = 17; 27.9%), None (n = 14; 23.0%), Other (n = 2; 3.3%)
Years of hacking experience (professionally)	<1 year (n = 9; 14.8%), 1–3 years (n = 30; 49.2%), 4–6 years (n = 11; 18.0%), 7–10 years (n = 4; 6.6%), None (n = 7; 11.5%)
Years of hacking experience (personally)	1–3 years (n = 29; 47.5%), 4–6 years (n = 20; 32.8%), 7–10 (n = 9; 14.8%), ≥10 years (n = 3; 4.9%)
Cybersecurity certification held	Yes (n = 49; 80.3%), No (n = 12; 19.7%)
Hackathon/CTF participation	1–5 events (n = 26; 42.6%), 6–10 events (n = 14; 23.0%), 11–15 events (n = 5; 8.2%), ≥16 events (n = 9; 14.8%), None (n = 7; 11.5%)

4.1 Certification Vs. Skills Test Performance (H1)

We compared the Cedar Bunny scores between those holding an active OSCP certification (N = 41) and those who did not (N = 108). A Mann-Whitney U test (U = 1096.00, Z = −4.79, p < 0.001; Fig. 2) revealed that OSCP holders (mean rank = 102.27) scored significantly higher than non-holders (mean rank = 64.65). This result supports H1, suggesting that OSCP holders performed better in the Cedar Bunny skill test than non-holders.

4.2 Certification Vs. Hacking Experiment Progress (H2)

Next, we examined the performance of the 61 participants who qualified for and completed the two-day experiment. We compared the hacking progress (scale 1–12) between OSCP holders (N = 25) and non-holders (N = 36). A Mann-Whitney U test (U = 253.00, Z = −3.13, p = 0.002; Fig. 3) indicated that OSCP holders (mean rank = 38.88) achieved significantly greater hacking progress compared to non-holders (mean rank = 25.53). This supports H2, suggesting OSCP holders still outperformed their non-certified counterparts.

Author Proof

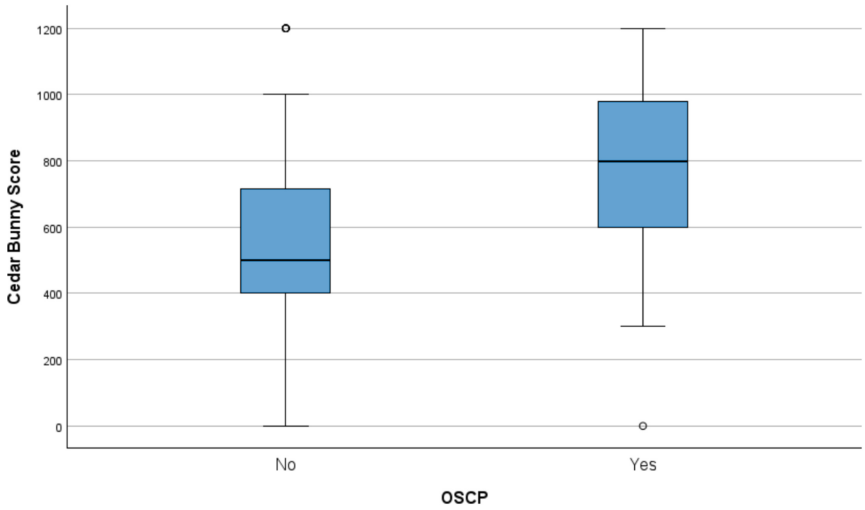


Fig. 2. Boxplot comparing Cedar Bunny skills test scores between participants holding an OSCP certification (N = 41) and non-holders (N = 108).

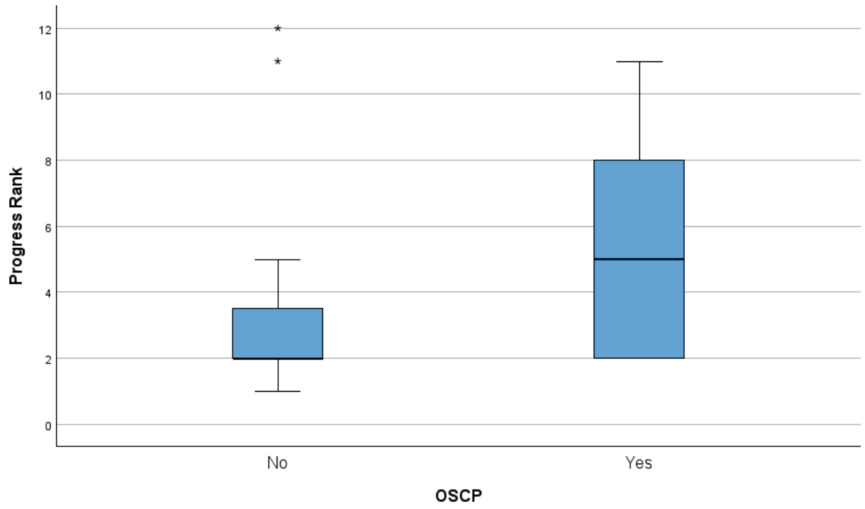


Fig. 3. Boxplot comparing Hacking Progress (1–12 scale) between participants holding an OSCP certification (N = 25) and non-holders (N = 36).

4.3 Skills Test Predictive Validity (H3)

We assessed the relationship between the skill test and experimental performance. A Spearman's rank correlation coefficient calculated for the 61 participants revealed a significant, moderate positive correlation between Cedar Bunny scores and hacking progress ($\rho = 0.344$, $p = 0.007$; Fig. 4), which supports H3.

Author Proof

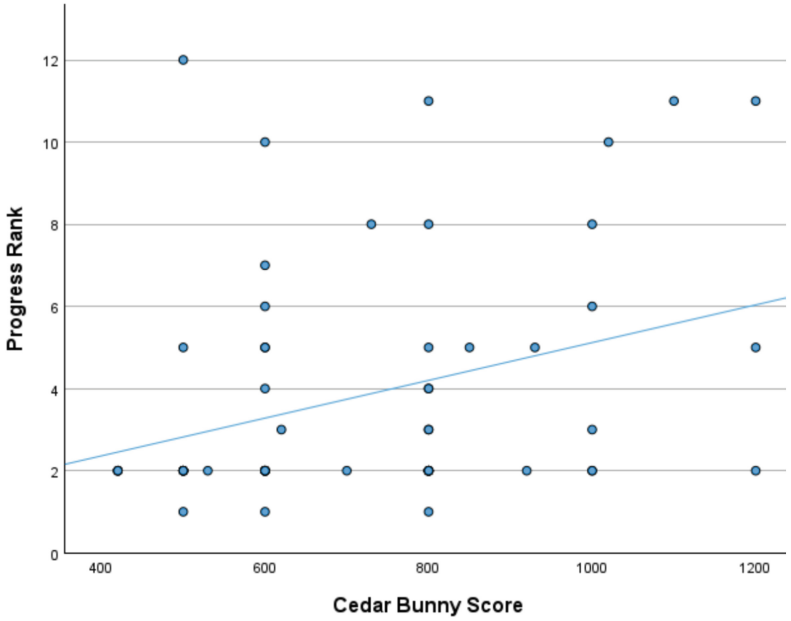


Fig. 4. Scatter plot showing the relationship between Cedar Bunny skills test scores and subsequent performance in the main hacking experiment (N = 61).

4.4 Skills Test Incremental Validity (H4)

To evaluate if the Cedar Bunny skills test provided unique predictive value beyond the OSCP certification, we conducted a hierarchical multiple regression (Table 3). In Step 1 that only included OSCP as the predictor, the model captured 12.8% variance ($R^2 = 0.128$, $F(1, 59) = 8.66$, $p = 0.005$). In Step 2 with inclusion of OSCP and Cedar Bunny scores, the model accounted for an additional 5.2% of the variance ($\Delta R^2 = 0.052$).

The standardized beta coefficients in the final model indicated that the predictive strength of the skills test ($\beta = 0.239$, $p = 0.060$) was marginal significant but comparable to that of the OSCP certification ($\beta = 0.285$, $p = 0.026$). This suggests that while certification remains a robust predictor, the skills test could capture similar variance in performance.

5 Discussion

The study result or confirmation of H1 indicates that OSCP is a meaningful credential and a valid predictor of hacking performance. Specifically, it is effective for qualifying participants who need to perform in the Kali Linux platform. The result supporting H2 indicates that the OSCP holders possess hacking skill beyond those who passed the threshold of a 420 Cedar Bunny scores. This suggests that the Cedar Bunny threshold score is likely a less demanding criterion than the OSCP certificate. An alternative explanation could be that the OSCP certification reflects on skills more relevant to our experimental tasks.

Author Proof

Table 3. Hierarchical Multiple Regression Analysis Predicting Hacking Progress from OSCP Certification and Cedar Bunny Skills Test Scores (N = 61).

Predictor	Model 1 (OSCP only)			Model 2 (OSCP + Cedar Bunny)		
	B	Std. Error	β	B	Std. Error	β
(Constant)	2.972	0.459		0.716	1.259	
OSCP	2.108**	0.716	0.358	1.679*	0.736	0.285
Cedar Bunny				0.003	0.002	0.239
R^2	0.128			0.180		
ΔR^2				0.052		
Model F	8.655**			6.363**		

* Significant at the .05 level
** Significant at the .01 level

The positive correlation between skills testing and performance confirmed H3, providing evidence that the Cedar Bunny skill test is a viable screening method. The hierarchical regression confirmed H4, illustrating that the Cedar Bunny test offered incremental predictive value beyond OSCP certification status ($\Delta R^2 = 5.2\%$), providing further evidence on the value of the Cedar Bunny skills test. However, the predictor of Cedar Bunny score was only marginally significant ($p = 0.060$), likely due to the statistical power of our sample size ($N = 61$). Nevertheless, the standardized coefficient for the predictor of Cedar Bunny score ($\beta = 0.239$) was comparable to that of the OSCP certification ($\beta = 0.285$). This similarity in effect size between Cedar Bunny scores and OSCP status suggests that the skills test can be a functional alternative to the certification. While the OSCP is an efficient fast filter, relying on it exclusively would systematically exclude competent participants who can also progress in the cyber range without the formal credential. The regression results revealed that the skills test successfully captured some additional variance demonstrating its practical utility and value for equitable recruitment.

5.1 Implications for Research and Practice

For researchers operating under strict time constraints, using OSCP as a primary inclusion criterion is a viable strategy for studies involving Kali Linux platform. The certification status can reduce the "false eligible" rate. In our screening process, OSCP holders passed the skills test at a much higher rate than non-holders, meaning that targeting this demographic would reduce the overhead of administering failed tests.

However, relying exclusively on credentials threatens the study's ecological validity and equity by introducing significant sampling bias. In our dataset, 36 of the 61 participants who demonstrated some of level of competence according to the skills test (59%) did not hold an OSCP. While H2 indicates that non-holders generally showed less hacking progress than their certified counterparts, the regression analysis suggests that skills test is practically valuable for recruitment. The skills test provided incremental predictive validity to OSCP and showed a predictive strength comparable to the certification.

Author Proof

These findings suggest a mixed screening approach may be optimal for participant recruitment in representative cybersecurity studies:

1. Fast-Track: Participants showing proof of domain-aligned performance certifications (like OSCP) from questionnaires could be consider sufficient proof of competence, by-passing any skills test and saving researcher resources.
2. Full-Screening: Participants without specific certifications should be offered the opportunity to complete skills assessment like Cedar Bunny.

This approach offers efficiency by reducing the testing volume while preserving equity by offering an alternative enrollment pathway for non-certified participants.

5.2 Threats to Validity

This study contains limitations that can be interpreted as threats to research validity and thus critical for future research to address:

- Construct Validity: A potential limitation lies in the measurement of "hacking progress." Though based on pre-defined flags, the hacking progress scores did not capture the stealth or efficiency of attacks that are important in professional settings. A participant who "smashed and grabbed" would score the same as one who moved stealthily, despite the latter demonstrating superior tradecraft.
- External Validity: The study was conducted in the SimSpace cyber range with the Kali Linux platform. Results may not generalize to other offensive environments, such as cloud security assessments or bespoke exploit development, where the specific tools taught in OSCP are less relevant. Furthermore, the Cedar Bunny skills test lacked validation evidence in the open literature, so its reliability and validity as a standardized instrument remain to be investigated.
- Internal Validity: A primary threat to internal validity is the confounding influence of professional experience. Participants holding an OSCP certification are more likely to be employed in active offensive security roles. Consequently, it is difficult to isolate whether their superior performance is attributable to the certification training itself or to the continuous practice of the skills in work settings. This confounding variable suggests that certification status may fundamentally overlap with job experience in professional populations.

6 Conclusion

This study demonstrates that while certifications like OSCP are highly efficient predictors of cybersecurity skills for Kali Linux, they should not be the sole gatekeeper for research participation. Skills-based testing for qualifying participants can help capture the "hidden talent" of the hacker community—those who possess the skills without the official credentials. By leveraging skills assessment in representative cyber environments in the absence of certifications, researchers can balance efficiency in recruitment and equity in representation, leading to more robust and generalizable results. Future research should examine a broader set of certifications, diverse hacking environments, and standardizations of skill tests to advance the proposed screening approach.

Author Proof

Acknowledgments. This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA) under Reimagining Security with Cyberpsychology-Informed Network Defenses (ReSCIND) program contract N66001-24-C-4504. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

Disclosure of Interests. None of the authors has any competing interests to declare that are relevant to the content of this article.

References

1. Allen, D., Herr, C.: Cyber Talent Identification and Assessment (2019)
2. Beltz, B., et al.: Guarding against malicious biased threats (GAMBiT) datasets: revealing cognitive bias in human-subjects red-team cyber range operations. Data Brief 112476 (2026)
3. Beltz, B., Chen, P.Y., Doty, J., Fonken, Y., Gurney, N., Hsing, H.W., et al.: Guarding against malicious biased threats (gambit): experimental design of cognitive sensors and triggers with behavioral impact analysis. arXiv preprint (2025). https://arxiv.org/abs/2512.00098
4. Butel, J.: Impact of Organizational Security Certification on Operational Security: Examining the Red Tape. MGA School of Computing Doctoral Papers (2023)
5. Crowne, D.P., Marlowe, D.: A new scale of social desirability independent of psychopathology. J. Consult. Psychol. **24**(4), 349–354 (1960)
6. Elzey, K., Jyotishi, S.R.: Combining degrees with quality certifications is a win for everyone. RealClear. Education. (2020)
7. Falah, A., Pan, L., Chen, F.: A quantitative approach to design special purpose systems to measure hacking skills. In: In: 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), pp. 54–61. IEEE (2018)
8. Giboney, J.S., Proudfoot, J.G., Goel, S., Valacich, J.S.: The security expertise assessment measure (SEAM): developing a scale for hacker expertise. Comput. Secur. **60**, 37–51 (2016)
9. ISC2: 2024 Cybersecurity Workforce Study. ISC2 Insights (2024)
10. James, J.E., Callen, J.: Cybersecurity certifications matter. Issues Inf. Syst. **19**(3) (2018)
11. Jansen, R.A., Rafferty, A.N., Griffiths, T.L.: A rational model of the Dunning-Kruger effect supports insensitivity to evidence in low performers. Nat. Hum. Behav. **5**(6), 756–763 (2021)
12. Knapp, K.J., Maurer, C., Plachkinova, M.: Maintaining a cybersecurity curriculum: professional certifications as valuable guidance. J. Inf. Syst. Educ. **28**(2), 101–114 (2017)
13. Kruger, J., Dunning, D.: Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J. Pers. Soc. Psychol. **77**(6), 1121–1134 (1999)
14. Lin, J. W., et al.: Comparing AI agents to cybersecurity professionals in real-world penetration testing. arXiv preprint (2025). https://arxiv.org/abs/2512.09882
15. Moneva, A., Ruiter, S., Meinsma, D.: Criminal expertise and hacking efficiency. Comput. Hum. Behav. **155**, 108180 (2024)
16. Offensive Security: OSCP Certification. https://www.offsec.com/pwk-oscp/. Accessed 23 Jan 2026
17. Paulhus, D.L.: Two-component models of socially desirable responding. J. Pers. Soc. Psychol. **46**(3), 598–609 (1984)

Author Proof

18. Ramezan, C.A.: Examining the cyber skills gap: an analysis of cybersecurity positions by sub-field. J. Inf. Syst. Educ. **34**(1), 94–105 (2023)
19. SANS Institute: Think like a hacker: Inside the minds and methods of modern adversaries. Bishop Fox (2022). https://www.sans.org/white-papers/think-like-a-hacker-inside-the-minds-and-methods-of-modern-adversaries. Accessed 23 Jan 2026
20. Siewertsen, R.: OSCP: a Journey to Become Certified in under 90 Days. Doctoral dissertation,. California State University San Marcos (2023)
21. SimSpace. https://simspace.com/. Accessed 12 Feb 2026
22. Soetan, S.A.: Design and Implementation of Reusable CTF Cybersecurity Recruitment Platform. Diva-Portal.org (2025)
23. Švábenský, V., Čeleda, P., Vykopal, J., Brišáková, S.: Cybersecurity knowledge and skills taught in capture the flag challenges. Comput. Secur. **102**, 102154 (2021)
24. Tay, A., Hayes, S.M., Wilson, D., Hall, E., Kaufman, D.: Gamified cybersecurity education through the lens of the information search process: an exploratory study of capture-the-flag competitions. Inform. Sci. Inf. Technol. **21** (2024)
25. Taylor, C., Arias, P., Klopchic, J., Matarazzo, C., Dube, E.: CTF: state-of-the-art and building the next generation. In: Proceedings of the 2017 USENIX Workshop on Advances in Security Education (2017)
26. Tran, B., Benson, K.C., Jonassen, L.: Integrating certifications into the cybersecurity college curriculum: the efficacy of education with certifications to increase the cybersecurity workforce. J. Cybersecurity Educ. Res. Pract. **2023**(2) (2023)
27. Yamin, M.M., Katt, B.: Inefficiencies in cyber-security exercises life-cycle: a position paper. In: AAAI Fall Symposium: ALEC, pp. 41–43 (2018)