The Simulated Security Assessment Ecosystem: Does Penetration Testing Need Standardisation?

William Knowles[a], Alistair Baron[a,*], Tim McGarr[b]

[a] *Security Lancaster, School of Computing and Communications, Lancaster University, Lancaster, LA1 4WA, United Kingdom*
[b] *British Standards Institution, 389 Chiswick High Road, London, W4 4AL, United Kingdom*

Abstract

Simulated security assessments (a collective term used here for penetration testing, vulnerability assessment, and related nomenclature) may need standardisation, but not in the commonly assumed manner of practical assessment methodologies. Instead, this study highlights market failures within the providing industry at the beginning and ending of engagements, which has left clients receiving ambiguous and inconsistent services. It is here, at the prior and subsequent phases of practical assessments that standardisation may serve the continuing professionalisation of the industry, and provide benefits not only to clients, but the practitioners involved in the provision of these services. These findings are based on the results of 54 stakeholder interviews with providers of services, clients, and coordinating bodies within the industry. The paper culminates with a framework for future advancement of the ecosystem, which includes three recommendations for standardisation.

Keywords: Penetration Testing, Security, Evaluation, Standards, Assessment

1. Introduction

In the presence of the seemingly inexorable increase in cyber attacks, how should organisations best pursue self-examination to accurately determine their resilience to such threats? One approach has been through the increase in services sold to these organisations that intend to replicate the methodologies and techniques (technical and social) of both internal and external malicious attackers, which are branded using a complex, and often confusing set of terminologies – something that we collectively describe here as "simulated security assessments". This paper seeks to explore the context in which these services are delivered, in order to determine best practices, and opportunities for further advancement.

The collective terminology of simulated security assessments uses the notion of simulation as it is established by Such et al. [1] in their definition of information assurance techniques. Simulation here is the practical imitation of threat actors within real-world environments, as opposed to the virtual alternative. Although the concept bears a strong relationship to non-contractual vulnerability research and its formally crowdsourced and contractual counterpart (e.g., bug bounty programs), the work presented within this paper is primarily concerned with contractual services procured from third party organisations.

Examples of such simulated security assessments include red team exercises, penetration tests, social engineering, and vulnerability scans. In practice, each of these services that can constitute a simulated security assessment has subtle differences, which are defined here for further context within this paper. The core feature of a vulnerability scan is its use of automatic tools; however, the use of automatic tools should arguably be followed up at minimum with a cursory manual review (e.g., false positive and negative verification), but at what level of analysis does manual review become a vulnerability assessment? Furthermore, how much manual review is required before an engagement can be considered a penetration test, or is such a label defined by the use of exploitation? This distinction is perhaps most notable and has been the most controversial within the security community. Labellings such as "IT Health Check" add a further dimension to this conundrum, as they exemplify the use of domain-specific labelling of types of simulated security assessment; in this case a form of penetration test. A long-term penetration test, possibly with greater scope of allowed activities (e.g., social engineering using approaches such as phishing or physical access) is often called a red team exercise, which attempts to test an organisation's cyber security capabilities against real-world simulations of persistent threats. Despite this wide variation in service offerings, the central motivation for their procurement is typically the same – to generate evidence (e.g., of the efficacy of security controls) that contributes as part of wider security risk management programs. Such evidence can then be converted into organisational-specific risks. The effectiveness of sim-

*Corresponding author
Email address: a.baron@lancaster.ac.uk (Alistair Baron)

ulated security assessments in generating this evidence for security risk management is exemplified in how their usage is no longer restricted to organisations pursuing the "extra mile" but in how it is rapidly becoming a mandatory requirement as part of many organisational standards.

The market for simulated security assessments is in a relative adolescence, and the complexity and ambiguity of service models (both in terms of service definitions and what is delivered against them) may be a consequence of this. Given this dynamic and the rapidly evolving nature of the market, it establishes the need to assess the current state of affairs, and to establish if there is a requirement for standardisation.

The research presented in this paper has been conducted in partnership with the British Standards Institution (BSI), the UK national standards body, which is responsible for originating many of the world's most commonly used management system standards, such as ISO 9001 and ISO/IEC 27001. The overall aim of the research was to assess the need for standardisation in the area of simulated security assessments, and provide guidance on what any proposed standards should include. We do not, however, restrict ourselves to recommending only formal standards, and utilise the broader definition of standardisation offered by De Vries [2]:

> "activity of establishing and recording a limited set of solutions to actual or potential matching problems, directed at benefits for the party or parties involved, balancing their needs and intending and expecting that these solutions will be repeatedly or continuously used, during a certain period, by a substantial number of the parties for whom they are meant."

The benefits of standardisation have been widely discussed, with links to increases in productivity, globalisation, exports, and general economic growth [3, 4, 5], as well as the facilitation of innovation [6]. Standards aid public procurement in decision making and risk management, but individuals involved with public procurement need to provide a greater input into the development of standards from the early stages [6]. Any recommendations for standardisation must be grounded in the realities of industry practices and client experiences, i.e. to establish the actual or potential problems to address. It is also important to understand links to existing standards, what level of self-organisation of the industry has occurred, and if this requires explicit standardisation. The research questions that this study will address are thus:

1. What standards currently exist for simulated security assessments?
2. What coordinating bodies exist within the industry for organisations and individuals?
3. How are current offerings perceived and what are the prevailing issues surrounding simulated security assessments?
4. Is there a need for additional standards, or to modify existing standards?

A requirement for action was identified which led to the publication of a preliminary white paper [7]. This paper provides an expanded and academically-focused extension to this work and makes the following contributions:

1. A review of the simulated security assessment ecosystem, detailing standards requiring assessments, standards for providing assessments, and individual qualifications for those that do.
2. An analysis of the performance of simulated security assessment providers over three phases of an engagement: pre-engagement, practical assessment, and post-engagement. Findings are based on the experiences gathered through 54 stakeholder interviews with providers, clients, and coordinators.
3. A framework for future advancement in the simulated security assessment ecosystem, which includes both standardisation and industry-led work activities.

This study predominately addresses the UK landscape, as this is where the stakeholder interviews were conducted. However, as will be evidenced in the remainder of the article, it can be argued that the UK is leading developments in this area. We will highlight examples of best practice in the UK that could be adopted internationally, as well as discussing international elements where appropriate. The recommendations for the advancement of the UK ecosystem should be broadly applicable in other countries.

The remainder of this paper is structured as follows. In Section 2 the academic literature concerning simulated security assessments is discussed. The methodology is outlined in Section 3. Section 4 then provides a review of coordinating bodies, standards, and qualifications within the simulated security assessment ecosystem. The simulated security assessment engagement process is then broken into phases in Section 5, which includes a discussion on stakeholder practices and experiences. Recommendations for future standardisation activities, along with industry-led improvement are described in Section 6. The paper is then concluded in Section 7.

2. Related Literature

Despite the significant body of academic literature that exists on specific techniques within technical security assessments, there has been limited focus on the wider ecosystem of simulated security assessment services. The literature that has attempted to address this domain falls into four broad areas.

2.1. Software Development Life Cycle

The first has emphasised the role of simulated security assessments within the Software or System Development

Life Cycle (SDLC), which has largely arisen through the U.S. Department of Homeland Security project "Build Security In". Software security best practices were discussed by McGraw [8, 9], emphasising the importance of integrating security into the SDLC, in particular the use of penetration testing. A similar emphasis was placed by van Wyk and McGraw [10] who outline various "touchpoints" (activities) in the SDLC to achieve this, including code review during implementation (a derivative "assurance technique" [11] of penetration testing), and penetration testing during configuration and deployment. A further study by Arkin et al. [12] focused specifically on software penetration testing, which emphasises the use of tools during the SDLC (e.g., static and dynamic analysis tools), along with the importance of contextualising assessments according to perceived risk posture.

2.2. Procurement

A second area of research has centred around the procurement of simulated security assessments. A high-level introduction to the motivations for doing so are provided by Hardy [13], while Bishop [14] provides discussion around the thought processes involved in scoping a meaningful penetration test, and addresses topics such as goal setting, attacker knowledge, resources, and ethics. Geer and Harthorne [15] highlight the contradictory drivers for penetration tests, which arise through clients desiring advertisable (in the context of demonstrating security to stakeholders) yet meaningful findings, and providers of services wanting to primarily succeed in the engagement's objectives (i.e., to discover flaws). The importance of penetration testing being only one part of a wider security risk management programme was stressed by Midian [16], while also introducing common issues found during assessments. The use of counterfactuals (i.e., "what-if" scenarios) in the process of security reasoning is explored by Herley and Pieters [17], and concludes that despite their challenges they are a "necessary evil". Penetration testing as a form of impersonation of such a counterfactual is discussed. Two further papers focus on procurement from the client's perspective, including establishing requirements for organisations providing such services. Tang [18] emphasises the importance of organisational standards, and proposes that procurers should look for companies certified by the Communications-Electronics Security Group (CESG[1]) IT Health Check Service (CHECK), the Council of Registered Ethical Security Testers (CREST), or ISO 17025 (for testing laboratories), and also have ISO/IEC 27001. The requirement for CREST certification is also proposed by Yeo [19]. Both CHECK and CREST are UK organisational certifications to provide independent assurance of simulated security assessment providers, and also offer individual qualifications. A high-level introduction to each, along with other standards and individual qualifications in the UK is provided by Xynos et al. [20]. This topic is examined in greater depth and supplemented with perceptions of stakeholders in this paper in Section 4.2.

2.3. Methodologies

The third area concerns research around the methodological characteristics of simulated security assessments. In most cases, methodologies have been established at a high-level of abstraction. An early paper by Pfleeger, Pfleeger and Theofanos [21] outlined such a methodology, while proposing breaking systems into objects that undergo transactions. A four element methodology (planning, discovery, exploit, reporting) was proposed by Bechtsoudis and Sklavos [22] who further describe a case study to identify common security issues and their implications. Goel and Mehtre [23] provide a basic introduction to the importance of vulnerability assessments and penetration testing, along with a linear methodology. Yeo [19] provides a more detailed methodology for the reconnaissance and attack phases of a penetration test, which is largely linear, with a cyclical element where compromise leads to the requirement for further enumeration. The modelling of penetration tests using petri nets has been explored by McDermott [24]. Thompson [25] provides discussion on penetration testing across three areas: building a test plan (e.g., in relation to a threat model), executing this plan (e.g., in terms of the types of testing, such as dependency testing), and the output of this process (e.g., having clear, detailed reproducible exploit scenarios). Geer and Harthorne [15] also discussed five aspects of application penetration testing: why (e.g., motivations for testing), who (e.g., assessor characteristics) what (e.g., the testing methodology), where (e.g., in terms of application subsystems), and when (e.g., at which point in the SDLC). Limited research has examined the application of such methodologies to niche scenarios. One exception to this can be found in the work on social engineering by Dimkov et al. [26], which proposes two methodologies: one which measures the environment surrounding a target asset, and does not involve social engineering the asset owner, and a further methodology which is more general, where the asset owner is in scope and unaware of the assessment.

2.4. Education, Training and Ethics

The fourth area revolves around ensuring the competencies of individuals performing simulated security assessments. A study by Guard et al. [27] assessed the characteristics of students most suited to conducting penetration tests within testbed environments. Skillset requirements and career development of penetration testers were discussed at a high-level by Caldwell [28]. Qualifications were explicitly referenced, which include three of the UK bodies, namely CHECK, CREST, and the Tigerscheme, along with the more internationally focused EC-Council Certified Ethical Hacker (CEH) and ISC^2 Certified Information

[1] CESG is the information security arm of the UK Government Communications Headquarters (GCHQ)

Systems Security Professional (CISSP). The challenges of designing information security courses that have potentially damaging consequences was examined by Logan and Clarkson [29] who identified a need for greater emphasis on the integration of ethics into courses. This need was also proposed by Saleem [30] and Pashel [31], both of whom provided a short analysis on teaching penetration testing, where the main emphasis was placed on preventing misuse of taught skillsets. At a high level, such ethical issues were discussed by Jamil and Khan [32] and Smith, Yurcik and Doss [33], with the latter highlighting the ethical dilemma created by a community which releases security assessment tools, and then also protects clients against individuals that use them. A similar theme can be found in Matwyshyn, Keromytis and Stolfo [34] who explored the ethics of security vulnerability research. Only three further papers provided a degree of detailed ethical analysis. The first was produced by Pierce, Jones and Warren [35] who established a conceptual model for ethics within penetration testing which contains five ethical "themes" of which integrity is at the core. The second was by Ashraf and Habaebi [36] who reviewed penetration testing ethics within the context of the Islamic faith. The third was by Mouton et al. [37] who examine the requirement for ethics specifically within the context of social engineering. Additional work has examined the ethical dimensions of conducting the most common type of social engineering attack, phishing. Finn and Jakobsson [38] present three experiments to measure susceptibility to phishing attacks, discussing how to make them realistic, and ethical. Although not specific to penetration testing, Reece and Stahl [39] have also conducted a UK-based stakeholder-led study to examine perceptions around the professionalisation of information security, which found mixed levels of support.

2.5. Summary

Four areas of academic research have been highlighted for simulated security assessments; however, two main criticisms can be applied to existing research. First, while the benefits of such assessments are widely espoused, there has yet to be a detailed review of the ecosystem for both the organisations providing such services, and the qualifications for individuals that work inside such organisations. Such standards and qualifications have been highlighted here where mentioned; however, in each case, references are largely cursory and a lack of analytical depth is evident. It does, however, highlight the dominance of UK standards and qualifications, which provides some validation of this paper's UK-centric focus. Second, there is an absence of any empirical review of the effectiveness of penetration testing in practice. Research has developed knowledge or established the need for knowledge in dynamic areas such as methodologies and ethics; however, we argue that future research requires a greater understanding of what occurs within real-world engagements.

3. Methodology

This paper presents the first comprehensive analysis of the simulated security assessment industry. The analysis focuses on UK-based services with reflections on the wider international implications. The study's methodology is visualised in Figure 1. The project can be seen to span three phases, which encompass the three research questions outlined in Section 1. Within Figure 1 the arrowed lines are used to denote contributing relationships; namely, between the sequential phases, the activities within these phases, and where the data sources originate that are used in their findings.

This first phase frames the contemporary *ecosystem* for simulated security assessments. This phase resulted in the comprehensive review of individual qualifications, organisational standards for those providing simulated security assessments, organisational standards that mandate an assessment, and the bodies that enforce them. Only through an understanding of what exists can the understanding of its performance be contextualised. This *performance* was the focus of phase two. To understand the realities of real-world engagements, in-depth interviews were conducted with 54 stakeholders about their experiences across three phases of an engagement: pre-engagement (e.g., scoping), practical assessment, and post-engagement (e.g., reporting). The simulated security assessment ecosystem is not a static entity; it is under continual evolution. Phase three is concerned with ensuring such *progression* occurs in a manner that benefits all industry stakeholders with the central objective of improved client security. The role of standardisation is considered here, along with opportunities for improvement by the industry itself. This article has been structured around the three phases: phase one (*ecosystem*) can be found in Section 4; phase two (*performance*) in Section 5; phase three (*progression*) in Section 6.

Two data sources were used within this paper: desk research and interviews. Each is described further within Section 3.1 and Section 3.2.

3.1. Desk Research

The predominant mode of data collection used within phase one was desk research, which established the foundations for later research within phases two and three. Data sources for desk research fall within four categories.

1. Academic literature was consulted to frame the research from the perspective of academia. A scarcity of existing literature was identified (see Section 2), which resulted in the majority of data being sourced from the remaining three sources.
2. Formal standards (e.g., those from ISO, IEC and BSI) were reviewed; in total, over 40, after a preliminary analysis to identify those that had relation to the ecosystem.
3. Consortia standards, which unlike formal standards that are published through international and national

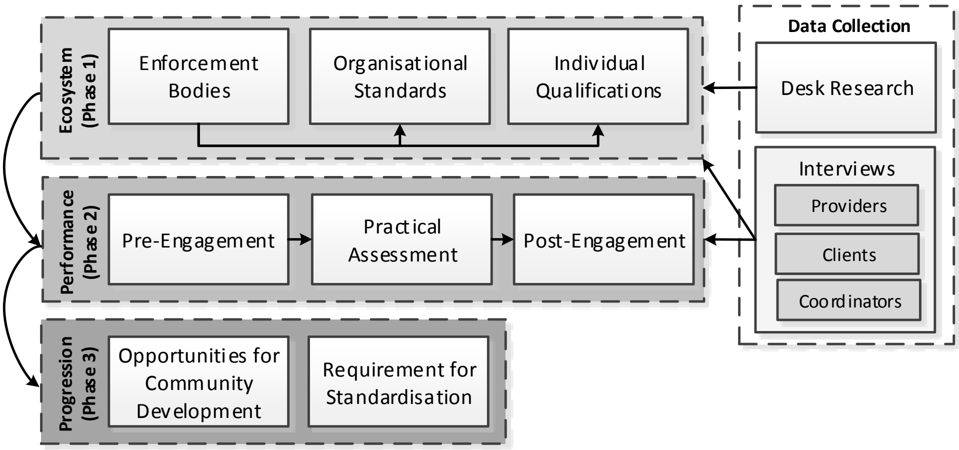

Figure 1: Methodology

standards bodies, required data to be collated through through public sources. This included the analysis of publicised information on websites (e.g., of trade associations) and specification documents (e.g., policies for membership requirements).

4. Community-led literature was also reviewed, which encompasses what has been produced by both individuals and collectives (e.g., technical reports by organisations and community standards).

3.2. Interviews

Stakeholder interviews were conducted to gather the perceptions and experiences of the challenges and opportunities present within real-world engagements and the wider ecosystem. Such interviews were primarily used within the analysis of phase two, but also informed phase one (e.g., for where information was not publicly available) and phase three (e.g., for perceptions of future industry direction). In total, 54 stakeholders were interviewed across 46 separate interviews. Stakeholders can be divided into three categories: those that provide assessments ("providers"), those that receive assessments ("clients"), and those bodies enforcing requirements on the other two stakeholder types ("coordinators"). The composition of each category is described below and visualised in Figure 2. The duration of provider interviews was between 40 minutes and 2 hours, for clients between 25 minutes and 1 hour 15 minutes, and for coordinators between 20 minutes and 1 hour 30 minutes.

Providers 32 stakeholders across 27 separate interviews, and 22 provider organisations. This includes providers of simulated security assessments in its various forms. All but 2 stakeholders were based in the UK. Out of 32 providers, 10 (across 8 organisations) were from CHECK accredited organisations, and 18 providers (across 13 organisations) were from CREST member companies. Insufficient information was collected to calculate the number of CHECK or CREST qualified individuals that were interviewed. Such individuals do not need to work for CHECK or CREST organisations, nor does one have to be CHECK or CREST qualified to work for one.

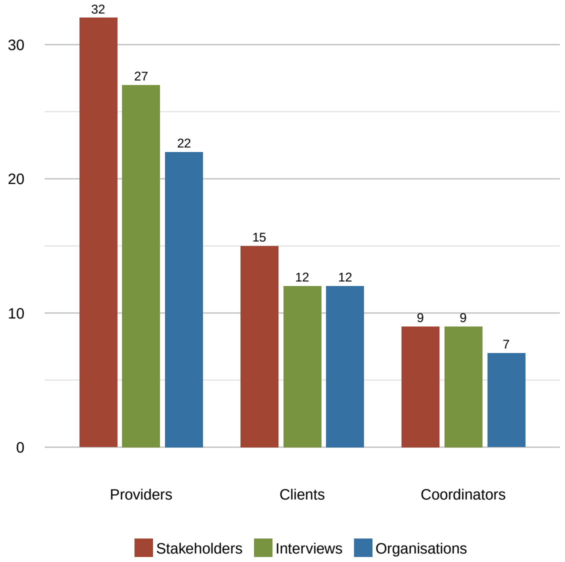

Figure 2: Stakeholder Composition

Clients 15 stakeholders across 12 separate interviews with 12 client organisations. To achieve a broad representation of client experiences, the organisational size of clients interviewed was highly varied. Client stakeholders ranged from micro enterprises (i.e., <10 employees) to large enterprises (i.e., >250 employees, with 3 stakeholders in organisations of >1000 employees, which included a financial institution). Furthermore, 5 representatives from UK local government were interviewed. Included within the total client count were 2 stakeholders who worked in consultancy roles (e.g. in one case, as a CESG Listed Advisor Scheme (CLAS) consultant) to procure penetration tests and identify remediation strategies for third parties.

Coordinators 9 stakeholders across 9 separate interviews with 7 coordinators. Included within this count were

2 stakeholders from provider organisations who also spoke about their roles within a coordinating body. This count of 9 stakeholders does not include the 3 providers who spoke about their work on the community standard, the Penetration Testing Execution Standard. Stakeholder organisations were: CESG; CREST; the British Standards Institution (BSI); the UK Department for Business, Innovation and Skills (BIS); Tigerscheme; Information Assurance for Small and Medium Enterprises (IASME); and Quality Guild (QG) Management Standards.

4. The Simulated Security Assessment Ecosystem

This section describes two perspectives on the simulated security assessment ecosystem. A client-focused perspective is taken in Section 4.1, which reviews standards that mandate particular forms of assessment. Section 4.2 then addresses this from the provider perspective in terms of qualifications, consortia/private standards, formal standards, community standards and methodologies for those delivering such assessments. Preliminary findings from the interviews for each section are discussed, with a more expansive analysis of their implications *in practice* given in Section 5.

4.1. Standards Requiring Simulated Security Assessments

There are a multitude of reasons why an organisation would procure or conduct a simulated security assessment. One notable driver arises through recommended or mandatory assessment requirements as part of a wider formal or private/consortia standard. The standards which make explicit reference to a requirement for some variation of a simulated security assessment are discussed below.

Cyber Essentials [40] is an entry-level organisational standard that provides basic assurance that an organisation is meeting minimum cyber security control requirements. It is targeted at private, not-for-profit and public organizations of all sizes, although it has particular relevance for small and medium enterprises (SMEs). It outlines two levels of certification: basic (no formal label) and "Plus". The Cyber Essentials standard requires the completion of a self-assessment form for basic certification, and self-assessment plus a third-party security assessment for Plus (including an external and internal vulnerability scan).

PCI DSS (The Payment Card Industry Data Security Standard) [41] enforces a business requirement for the information security of organisations handling payment transactions, including those by credit and debit card. Compliance with PCI DSS is not a legal requirement (with certain geographical exceptions) but instead a requirement enforced through business terms (e.g., non-compliance can result in penalty fines). Requirement 11.2 mandates quarterly vulnerability scans, while requirement 11.3 mandates penetration tests at least once per year and with any signifiant infrastructure or application modification.

ISO/IEC 27001 – an Information Security Management System (ISMS) standard – can be contributed to with simulated security assessments as audit evidence. In this case, such assessments are encapsulated as a security control under the umbrella of a "technical compliance review". Security controls within ISO/IEC 27001 are not mandatory (organisations opt-in or opt-out of security controls based upon a risk assessment), and therefore those under audit to pursue such certification are under no obligation to possess audit evidence of the results of a simulated security assessment. ISO/IEC 27001 is, however, widely used as the basis for other assurance schemes, and in some cases the "technical compliance review" becomes a mandatory requirement (potentially along with other security controls). One such example is the CESG Assured Service Telecoms CAS(T)[2] for telecommunication environments.

IT Health Checks (i.e., CHECK assessments) are another government standard requiring some form of simulated security assessment. IT Health Checks are for UK public sector bodies (including local governments) who wish to participate within the network that interconnects them: the Public Services Network (PSN).

ISO/IEC 15408 – more commonly known as the Common Criteria – outlines the requirements for the "secure functionality of IT products and for assurance measures applied to these IT products during a security evaluation" [42]. Penetration testing is frequently cited within the vulnerability assessment requirements of ISO/IEC 15408-3:2008 [43].

Interview Findings of Standards for Clients

Two important survey findings were made regarding the use of simulated security assessments for compliance where security controls are established.

Firstly, stakeholders from all three categories felt the link between simulated security assessments and **ISO/IEC 27001** was currently "disparate" and poorly documented. Two interrelated approaches for establishing a link emerged from early interviews, and subsequent stakeholder views were widely positive for both. The first approach was to establish a clear link between the activities within a simulated security assessments and ISO/IEC 27002 security controls, and the second was to establish greater auditor guidance for using assessment findings as audit evidence, within the larger ISMS audit. Arguably, the former must happen to enable the latter. Criticism was expressed by one stakeholder, whose views are notable due to their proximity to the standardisation process. This stakeholder felt the approach was at odds with the ISO/IEC 27001 model, which was not about security in itself, but knowing insecurity and planning for continuous improvement. This stakeholder added: "Why favour a particular method over another? Why is penetration testing better than auditing

[2] https://www.cesg.gov.uk/servicecatalogue/service_assurance/CAS/Pages/CAS.aspx

security records?" Risk must be identified to be managed, however, and for other stakeholders, the enthusiasm was focused on the ability of simulated security assessments to assess controls in demonstrable terms.

Secondly, whilst the motivations behind **Cyber Essentials** were widely applauded, it was criticised for its lack of target market, while provoking explicit and widespread confusion about its implementation. Such confusion arose primarily from the heterogeneous approaches of the accreditation bodies. Frequent remarks concerned the integration of companion standards within Cyber Essentials, where vulnerability assessments were required (or where they were not), and the separation of accreditation and certification status. Some providers further questioned whether consistency could be achieved due to the ambiguity in the testing guidelines, and the subjectivity required to implement them.

4.2. Standards and Qualifications for Providing Simulated Security Assessments

Competence requirements can be established at both the organisational and individual level. This section provides a discussion of the current state of the market for both, along with the views of stakeholders on the requirement for modified or new standards in these areas.

4.2.1. For Individuals

Budding and established professionals are now faced with a multitude of choices for qualifications across a range of skill levels and topic scopes. UK qualifications for simulated security assessments primarily arise from four providers: CESG, CREST, Tigerscheme[3] and the Cyber Scheme[4].

CESG has established a qualification scheme for the IT Health Check Service (CHECK), which has been in operation for over a decade. The two levels of CHECK qualification are the **CHECK Team Member** and the **CHECK Team Leader**, the latter of which is split over two qualifications for infrastructure and web applications. The current format requires candidates to have obtained a certain type and level of industry qualification and Security Clearance (SC) to allow them to handle sensitive data, amongst other publicly undisclosed factors. The three remaining qualification bodies, CREST, Tigerscheme and the Cyber Scheme, provide the industry qualification. Their content, level and equivalence to CHECK qualifications are shown in Table 1.

CREST has since emerged as the predominant industry-led professional qualification body within the UK, and its qualifications can be seen to span four tiers. In order of required proficiency, they are CREST Practitioner (requiring an estimated 2,500 hours of experience), Registered Tester (6,000 hours; CRT), Certified Tester (10,000 hours; CCT) and Simulated Targeted Attack and Response (STAR). It is at the Certified tier that specialism occurs in the areas of infrastructure or web application security. STAR is a framework created to provide intelligence-driven red team penetration tests for the critical infrastructure sectors. Currently the main implementation of STAR is CBEST[5], which specifically targets the financial sector STAR qualifications ensure that competency requirements are met to perform such engagements. Two forms of STAR qualifications exist: those for managers (i.e. those who lead STAR teams) and those for technical specialists. The **Tigerscheme** and **Cyber Scheme** qualifications follow a similar structure to the lower three CREST qualifications, each with a beginner, intermediate and advanced qualification. An equivalent to CREST's STAR qualifications is not available from other qualification bodies.

CESG has also launched a separate scheme, which forms a competence framework that is described as a certification rather than a qualification: the **CESG Certified Professional** (CCP). The CCP is a framework that defines seven roles, one of which is "Penetration Tester". Each role has differing levels of competence, which are aligned with the responsibility levels defined by the Skills Framework for the Information Age (SFIA)[6] and the skill levels defined by the Institute of Information Security Professionals (IISP). Four levels are defined for the Penetration Tester role: SFIA Responsibility Level 3, 4, 5 and 6. CCP, while listed in Table 1, does not currently contribute to a CHECK qualification assessment.

The list of qualifications described is not exhaustive and is UK focused, which has meant many non-UK training courses and qualifications have been omitted; this is a consequence of interview findings, which found a greater emphasis on UK qualifications for recruitment. Despite this, there are two qualifications worthy of note, both of which are from US-based providers. First, the International Council of Electronic Commerce Consultants" (EC-Council) **Certified Ethical Hacker** (CEH). This qualification can be positioned at the lower spectrum of the entry-level classification. Despite such positioning, it was identified to be frequently cited within job advertisements (typically supplementary to those of Table 1) and it has seen integration into some UK academic courses. CEH is not assessed through virtual lab examination, however, and uses only multiple-choice examination. Second, the **Offensive Security Certified Professional** (OSCP). In contrast to CEH, the interviewees perceived OSCP to be increasingly popular within the UK market for recruitment due to its rigorous technical virtual lab examination, and is the predominant requirement for US-based organisations. Since the conclusion of this interview process, CREST has partnered with Offensive Security to establish equivalency between OSCP and CRT[7]. This allows holders of OSCP

[3] http://www.tigerscheme.org

[4] http://www.thecyberscheme.com

[5] http://www.bankofengland.co.uk/financialstability/fsc/Pages/cbest.aspx

[6] http://www.sfia-online.org

[7] http://www.crest-approved.org/

Level	Qualification Bodies and Qualifications				
	CESG		CREST	Tigerscheme	Cyber Scheme
	CHECK	CCP			
Entry	N/A	SFIA Responsibility Level 3	Practitioner (CPSA)	AST	CSA
Intermediate	Team Member	Level 4	Registered (CRT)	QSTM	CSTM
Advanced	Team Leader	Level 5	Certified (CCT)	SST	CSTL
		Level 6			
Red Team	N/A	N/A	STAR (CCSAM and CCSAS)	N/A	N/A

Table 1: Penetration Testing Qualifications

to obtain CRT subject to certain stipulations (i.e., a fee and a multiple-choice and long-form examination within six months). CRT obtained in this manner, however, cannot be used as part of the CHECK application process.

Interview Findings for Individual Qualifications

The consensus amongst stakeholders was a strong opposition to any form of new standard for individuals. Opposition was twofold. Firstly, the techniques and skills used evolve at a rapid pace, which would be infeasible to capture and keep current within a standards type document. Secondly, while the current system is not without fault, existing consortia providers within the UK have done an exemplary job of raising, setting and assessing the competence of individuals that conduct simulated security assessments, and furthermore, the UK is ahead of the rest of the world in this regard.

Many stakeholders, however, did feel that there was a growing need for an independent body, modelled in the same vein as medicine or law, in order to continue the professionalisation of simulated security assessments. Taking medicine as an example, key indicators of its professionalisation are the internationally recognised standards for its practices, and regulation bodies for individuals with powers such as being able to revoke the right to practise. Some providers felt that standards bodies, such as BSI, could facilitate the internationalisation process by working with technical assessors such as CREST and Tigerscheme. However, not all stakeholders were positive about such an endeavour. It was noticeable that those supportive were predominantly in positions of management, whose natural proclivity is one of control. Those engaging in the practical elements of security assessments, the practitioners, had fears about the potential future exploitation of such a scheme to regulate those who wish to conduct cyber security research. It was felt that such a situation would negatively impact the industry as a whole, and lead to the loss of the UK's competitive advantage.

4.2.2. For Organisations

Confidence in a provider's process readiness to deliver simulated security assessments can be created through organisational standards. Such standards fall broadly into two areas: those from consortia and those from formal standards organisations (e.g. BSI directly or ISO/IEC). Private/consortia standards are those used most predominantly; such standards can be visualised using a tiered model, in which the hierarchy is formed by the level of *intended* rigour of the simulated security assessments that they provide. The reader should note that this is a rough categorisation, and there remains potential for offerings to move between tiers based upon client requirements. This tiered model is presented in Figure 3.

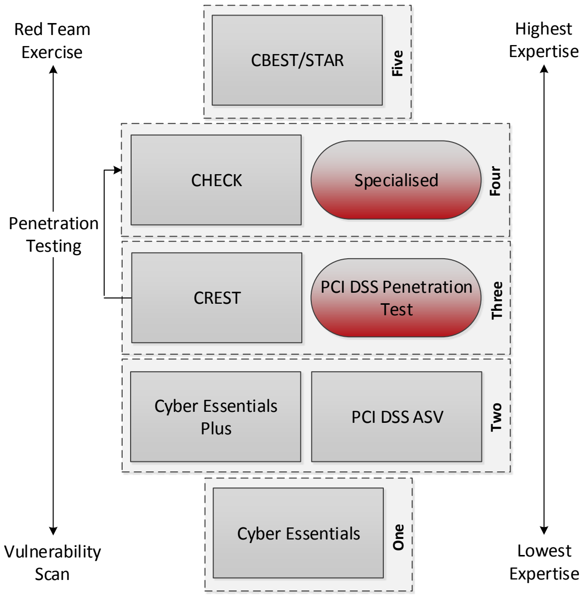

Figure 3: A Tiered Model of Provider Standards

Tier One is concerned with external vulnerability assessments. The **Cyber Essentials** scheme [40] was introduced in Section 4.1, along with its two types of certification: Basic and Plus. Provider organisations can be

professional-qualifications/oscp-and-crt-equivalency/index.html accredited (i.e., to become certification bodies) to deliver one or both levels by one of the four accreditation bodies.[8] Only one accreditation body, CREST, has implemented their version of the Cyber Essentials scheme to deliver simulated security assessments at this tier. CREST mandates that external vulnerability assessments must be conducted, on top of the "core" requirements of the scheme (i.e., a self-assessment form), in order to ensure border controls are properly implemented [44].

Tier Two expands the engagement scope of Tier One vulnerability assessments. Two main standards exist. The first is **Cyber Essentials Plus**. All accreditation bodies deliver assessments that meet the criteria set forth by the "Common Test Specification" [45] which outlines the mandatory internal and external assessments, along with success criteria. Assessments include vulnerability scans and configuration reviews (e.g., assessing ingress malware filtering at the boundary, server, and workstation levels). The second fulfils requirement 11.2 of **PCI DSS** for periodic vulnerability scans. The PCI Security Standards Council mandates that such vulnerability scans must be conducted by an Approved Scanning Vendor (ASV). Certification to become an ASV involves a simulated assessment on PCI Security Standards Council infrastructure to evaluate technical competence, and organisations must recertify annually.

Tier Three sees a shift towards increasingly adversarial engagements, and what are widely considered *penetration tests.* The most well-established industry-led body providing certification here is **CREST** The application process covers four domains of organisational capability: (a) information security; (b) operating procedures and standards; (c) methodology; and (d) personnel security, training and development. Both the ISO/IEC 27001 and ISO 9001 management system standards are referenced in CREST's guidance for applicants, but not mandated. However, CREST does require evidence of operational commitment to the implementation of an information security management system (ISMS) and a quality management system (QMS). Furthermore, CREST requires a clear and documented complaints procedure for Member Companies, with an escalation path that makes direct reference to the CREST complaints process for independent arbitration. **PCI DSS** requirement 11.3 for periodic penetration tests does not mandate any certification requirements for providing organisations. Instead, a de facto standardisation is enforced (represented by the red area within Figure 3) through the certification process when an appointed individual and/or organisation assesses whether requirements have been met to an appropriate level To facilitate this process PCI DSS have released supplementary penetration testing guidance [46], which includes recommended provider competencies (e.g. qualifications), methodologies (notably including emphasis on the importance of exploitation), and reporting. Within competency guidelines, it is noteworthy that the only organisational certifications promoted were the UK's CHECK and CREST.

Tier Four bears a close resemblance to Tier Three with respect to assessment expectations and required competencies; however, it differs in its establishment of "non-standard" requirements (e.g., for providers to have achieved security clearance). The primary standard here is **CHECK**, a governmental initiative operated by CESG, which enables approved organisations to perform penetration tests for government departments and public sector bodies (including Her Majesty's Government). Organisational approval requires evidence submission in two areas: (a) capability (e.g. testing methodology and examples of previous reports) and (b) the composition of a CHECK team (at least one CHECK Team Leader and one CHECK Team Member). There are a multitude of **specialised** engagement types that would fall into this category but have no formally defined offering. An example would be penetration tests involving safety-critical infrastructures (e.g., Industrial Control Systems). The requirements present within this level may also apply to some CREST engagements.

Tier Five engagements are those that require a similar or higher level of expertise as Tier Four, but differ predominantly in the length of the engagement and list of permissible activities. Tier Five may therefore be considered a form of *red team engagement.* Although many providers have red team capabilities there currently only exists one organisational standard to provide oversight within this market. This falls under what CREST refer to as the **STAR** framework, which offers threat intelligence-led red team exercises to the critical infrastructure sectors. Currently the main implementation of STAR is **CBEST** [47, 48] which establishes mandatory testing requirements in the financial sector. CBEST engagements may utilise government issued intelligence provided by the UK Financial Authorities, which may not be available for other STAR engagements.

Formal Standards: The discussion has focused on the private/consortia standards that dominate this domain. However, formal (technical) standards also exist that describe activities relating to simulated security assessments, or mandate their use. One widely used standard for security evaluations is ISO/IEC 15408 (Common Criteria). The methodology for such evaluations are outlined in ISO/IEC 18045. Various challenges prevent the widespread application of ISO/IEC 15408 to simulated security assessments, such as high information requirements about target environments, and the challenges of applying it to a dynamic and live system. However, attempts to mitigate this have been provided in supplemental standards; for example, PD ISO/IEC TR 19791 extends ISO/IEC 15408 to cover operational environments, while ISO/IEC TR 20004 uses the Common Weakness Enumeration (CWE) and the Common Attack Pattern

[8] This can be seen through the example of provider organisations accredited by CREST: `http://www.cyberessentials.org/certifying-bodies/`

Enumeration and Classification (CAPEC) frameworks to support ISO/IEC 18045 vulnerability assessments.

Interview Findings of Standards for Providers

Isolated criticisms were raised against consortia/private standards (e.g. a lack of independence from industry in their governance); however, the predominant voice amongst stakeholders was that they have done much to raise the standard of operational readiness and professionalism of providers within the UK. Indeed, approval for both **CREST** and **CHECK** was frequently cited as a motivation for the adoption of management system standards by providers, predominantly ISO/IEC 27001 and ISO 9001.

Although the benefits of some formal standards were espoused (e.g. the thoroughness of Common Criteria), the consensus was that it would prove difficult to implement these standards for the types of services and timescales for testing that clients were demanding. A small number of providers suggested standardising a methodology; however, this was mostly only seen as an option if standardisation was forced. Other stakeholders questioned the benefit of such a "high-level" standard, feeling there was already a significant quantity of information in the public domain on this topic. Efforts to create such a standard have been considered (and continue to be) by the subcommittee that developed the ISO/IEC 27000 series (ISO/IEC JTC 1/SC 27), although there is not yet any standard published or in development on this topic.

The performance of these standards and certifications for providers *in practice* (including methodologies used) will be discussed in Section 5.

4.3. Community Standards and Methodologies

Many of the guidelines and standards for conducting assessments have not come from formal standards institutions, but instead have been generated by the security community and other interested stakeholders. This section details those activities self-described as standards, and then finally those that can be broadly defined as guidelines and methodologies.

The Penetration Testing Execution Standard (PTES)[9] is a community standard that provides guidelines for the full scope of penetration testing activities over seven stages: pre-engagement, intelligence gathering, threat modelling, vulnerability analysis, exploitation, post-exploitation, reporting. Although the majority of these stages are technical evaluations, PTES itself does not specify technical guidelines on how to conduct a penetration test engagement, instead describing the process at a conceptual level. However, PTES has produced a set of technical guidelines to accompany the standard[10], which includes the specification of particular tools and instructions on their use. A further community standard was produced by the Open Web Application Security Project (OWASP) project, whose aims are to improve the state of web application security through the provision of guidelines, reports, and tools. OWASP have created the Application Security Verification Standard (ASVS) v2.0 which outlines a methodology for assessing web application security controls.

Other publications are available that do not purport to be standards, but instead act as guidelines and methodologies. Some are generically focused including the National Institute of Standards and Technology (NIST) Special Publication 800-115 [49] (which is considered best practice in PCI DSS [41]) and the Open Source Security Testing Methodology Manual (OSSTMM) [50]. Other guidelines and methodologies are more technology-specific, such as OWASP's "Testing Guide" for assessing web applications. Furthermore, there have been industry-specific publications that address the challenges of assessing systems within environments that require non-standard approaches during a simulated security assessment. For example, the United States Department of Homeland Security have produced high-level guidelines on the use and challenges of penetration tests for assessing Industrial Control System (ICS) security [51]. A more thorough methodology is presented by the National Electric Sector Cybersecurity Organization Resource (NESCOR) [52] which addresses penetration tests for assessing electric utilities in particular. NESCOR includes within its scope guidelines for assessing embedded components, which is largely unaddressed within other methodologies in favour of network and web application security. An extension of this can be seen in the Advanced Metering Infrastructure (i.e., a sub-system within the smart grid) Attack Methodology [53]. This publication outlines a methodology at the technical level for penetration testing embedded devices that exist outside of a utility's security perimeter (e.g., on customer premises).

Interview Findings for Community Standards and Methodologies

Due to the pragmatic nature of the publications described within this section, their impact will not be discussed here, but rather in Section 5, in order to frame this discussion within the context of experiences from real-world engagements.

4.4. Summary of key findings

Several key findings on the existing ecosystem and how it is viewed by stakeholders are worth highlighting:

- The link between simulated security assessments and ISO/IEC 27001 should be strengthened.
- The UK leads the rest of the world in raising, setting and assessing the competence of professionals conducting simulated security assessments.
- There is support for an independent body, similar to to medicine or law, to continue the professionalisation of the field.

[9] `http://www.pentest-standard.org/index.php/Main_Page`

[10] `http://www.pentest-standard.org/index.php/PTES_Technical_Guidelines`

- Existing consortia/private standards have done much to raise the level of operational readiness and professionalism of providers within the UK, including the adoption by providers of management system standards (e.g. ISO/IEC 27001 and ISO 9001).

5. The Engagement Process

Stakeholders were also questioned on their practices and experiences around simulated security assessments. To contextualise the findings, the terminology for an engagement's subprocesses has been defined through a reference model in Figure 4. This reference model was derived from stakeholder responses. The model splits an engagement into three broad phases: pre-engagement, practical and post-engagement.

The pre-engagement phase is concerned with establishing the parameters of allowed activity for the practical assessment. There will be some form of initial interaction which may be initiated by the provider or the client. A chosen methodology (e.g. questionnaires or interviews) will be used by the provider to generate a scoping proposal. This proposal may go through multiple rounds of negotiation. The client will then sign off on the proposal before the practical assessment begins.

The practical assessment phase involves the exposure of a client system or component to a simulated attack. The phase begins with information gathering. This may uncover systems that necessitate further discussions around the engagement scope (e.g. if a client uses systems owned or operated by third parties and there are questions of testing authorisation). A provider may conduct a threat analysis or move straight to its subsequent stage, a vulnerability analysis. Exploitation of identified vulnerabilities may occur in order to attempt penetration of the system, and to gain access to additional resources (e.g. sensitive data or higher system privileges). The subprocesses of the practical assessment stage may go through multiple repetitions (e.g. a compromised system may be connected to another internal network which, if under scope, can also be attacked).

The post-engagement phase is concerned with the delivery of findings to the client, usually in the form of a written report. The majority of providers will supplement this with additional forms of client interaction (e.g. final meetings) in order to educate them about the findings and the remedial actions that need to be undertaken.

Comments resulting from the stakeholder interviews concerning these three stages of the penetration test engagement will now be discussed in turn.

5.1. Pre-Engagement

> *The quality of the marketing collateral of penetration testing companies leaves a lot to be desired. I think it's a marketplace that's shrouded in mystery and myth. It's very difficult as a person wishing to purchase penetration testing and IT Health Check services ... to assess the marketplace and find out whether or not your potential vendors will satisfy what you require, other than them being able to say that they're CREST or CHECK registered ... it almost feels like you need to be an expert yourself to buy an expert to come in and help you ... Being able to come up with a framework with which you can engage these suppliers, and understand the nature of the different tests that they will do, and how they will treat that information in terms of reporting it back, and there being some consistency across the marketplace ... I think that would be a very welcome development.*
>
> **A client of penetration tests**

5.1.1. Terminology

There was a notable sense of confusion and frustration amongst stakeholders about the ambiguity in what constitutes a penetration testing service. Such ambiguity was evident from the varied service definitions of providers, in particular around the level of exploitation that occurs during engagements. A number of providers stated that vulnerabilities within engagements were not exploited by default, with additional value provided through theorised exploitation and/or false negative and positive verification. This caused a commonly cited issue during the tender process, which was found to be increasingly common for the procurement of simulated security assessments. Clients were often found to be unable to differentiate between providers, even amongst some of those that had approved CHECK or CREST status, while providers argued that clients often failed to understand their requirements, provided limited opportunities for consultation, and made procurement decisions based predominantly on economic factors. Providers argued that this could lead to clients failing to procure a level of testing rigour appropriate to the requirements of their environment. Some providers felt this was in part because clients are not concerned with the quality of the test: "clients are just looking for a tick in the box and resent any issues found". The issue from client and provider perspectives is related, and can arguably be reduced to issues with terminology for defining services.

The definition of consistent terminology was widely supported amongst providers (18) and clients (5)[11]. Another two providers expressed support at a conceptual level, but argued practical definitions were difficult to determine; the subjective nature of exploitation was cited as an issue by one provider. Questions around terminology arose from an early interview with a provider who suggested that the market would benefit from BSI working with industry partners to define testing types. This

[11] These figures were reached without all stakeholders being questioned on the topic due to time restrictions.

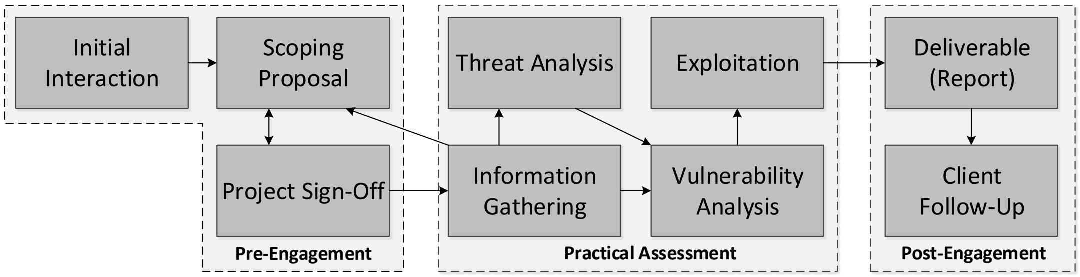

Figure 4: Phases of a Penetration Testing Engagement

provider argued that "it might not be right; it might get slaughtered, but it's a stake in the ground", where clients can say they want a peer defined simulated security evaluation (e.g. vulnerability assessment or penetration test) and have a clearer understanding of the service that they desire and what will be delivered.

As the support for terminology definitions suggests, the industry is acutely aware of the issues caused by the lack of precise terminology. One initiative with potential industry impact is the community standard, PTES. The upcoming version of the PTES is stated to have "levels" of testing. Supportive providers felt levels would empower clients and facilitate the process of procuring a certain type or level of test. If a provider was then to fail to deliver the requirements of that level, they would be in breach of contract. One non-PTES provider was supportive of a level approach; however, they urged caution with definitions, as part of the process of adding value is having the power to deviate. Such issues could easily be addressed through clarifying testing requirements in pre-engagement negotiations. If a standard is too specific, however, it could cause issues if clients do not need an aspect of that test, and do not understand why they do not need it. Providers would then need to deliver unnecessary services to meet that level.

An alternative solution proposed by one coordinator used a measure of the client's risk appetite to map onto industry services; however, it received strong opposition due to the difficulty in computing risk appetite (even amongst those versed in its specialism) and the lack of potential for internationalisation where many providers wished to focus their efforts.

5.1.2. Scoping

The scoping procedures of providers were found to have formed a de facto standard, with strong commonalities in the basic stages (see Figure 4), methodologies to derive client requirements (predominantly questionnaires), the structure of scoping proposals, and types of testing proposed (almost wholly white box). Client views on scoping were polemic. Providers were largely seen as providing adequate assistance; however, in some cases they were criticised for their excessive use of questionnaires and lack of face-to-face meetings, even by larger clients. One CESG Listed Advisors Scheme (CLAS) consultant argued that you "often can't interact with penetration testing providers". For some providers, especially the micro enterprises, face-to-face meetings were pursued at all opportunities to differentiate themselves against the rise of "faceless" providers.

Larger enterprises were considered to be the only clients who understood their requirements, and this often manifested in engagements with strict guidelines, goals and deliverables. Small providers were deemed to require greater levels of assistance, but have grown increasingly knowledgeable over the past three to five years through their periodic audits. A common area of contention amongst stakeholders was found in industries that mandate simulated security assessments; notably in the CHECK scheme for Public Services Network (PSN) compliance and PCI DSS. Multiple CHECK requiring clients expressed the opinion that their peers were intentionally narrowing the scope of engagements to minimise risk for any issues found due to the punitive nature of the scheme. Two CHECK providers stated that they had heard of such issues themselves. One client insisted that they were interested in having an expansive scope for the CHECK scheme. Such an approach provides financial benefits over having one approach for aspects critical to PSN compliance, and another non-mandatory test for other services. However, the punitive nature discourages such an approach, with the additional complexity that any issues found lead to poor reflection of security capabilities compared to peer organisations. The Cabinet Office does provide a four-page document (two pages of content) on scoping CHECK; however, for the clients the call was clear: greater guidance is needed to ensure consistency within the scheme.

5.1.3. Authorisation

Questions of authorisation arose when engagements involved third party services. All providers stated confirmation was sought before engagements began; however, the methods used varied. The preferred method involved the client signing a document to state their legal authority to test systems within scope, with the provider requiring no further proof. One provider stated this was because it was "too time-consuming to check it all". A minority of providers required explicit confirmation from the third party. Cloud services were an exception, with providers often demanding email or written confirmation from the cloud service. Such authorisation was found to be obtained with relative ease, except for smaller or "non-Western" cloud services. Providers stated that undisclosed third party systems were often uncovered during the initial reconnaissance stage of the practical assessment, notably with mail servers, and that the lack of third party authorisation was a common reason for delayed testing.

5.1.4. International Engagements

Providers were questioned on their understanding of the legality for conducting engagements outside UK borders. Providers were largely unaware, with the bigger providers stating that such engagements would be cleared by their legal department. The general approach was to offset risk onto the client on the assumption they would have greater knowledge of local laws, and to ensure to never stray from the scope set out for the engagement. Legal cover would then be provided by authorisation from the client. One provider felt there was not enough legal guidance around the Computer Misuse Act (1990), even within the UK. "Where does the Misuse Act stop and a new law begin?" This proved to be a bigger issue for smaller provider organisations. One such provider stated that they have had multiple enquiries about work from the USA; however, they have not taken on business because they do not understand how "protected they are". This provider felt that UK government Trade & Investment department should be making more effort within this domain. "They have to understand that process and advise if they're wanting to push an export strategy" for cyber security services.

5.2. Practical Assessment

Known hostilities towards standardisation of the technical aspects of simulated security assessments led to a strategic choice in study design to focus on other aspects of engagements; however, stakeholder interviews identified three areas of note.

5.2.1. Exploitation

The first bears relation to the question of terminology, and concerns the extent of provider exploitation, which can be seen as a key differentiator for service definitions. For many providers, there was an aversion to exploitation. For some providers their default policy was to not exploit, beyond basic false positive and false negative testing (e.g., for SQL injection, using a single apostrophe to raise a database error). In one case the provider expanded upon their approach, describing the creation of scenarios to determine if an attack is feasible "on the balance of probability". Furthermore, one provider stated that while exploitation occurred on some of their engagements, "most" of their clients do not like it, and so it is not conducted. Where exploitation was described as being used, the general consensus amongst providers was to "only use exploits we're sure about", "talk to clients if there is a risk to live services" and to liaise with the client and "seek approval" before exploitation. One provider described how sometimes clients will not want the provider to exploit, but would want them to continue on as if exploitation has worked. In such a scenario, the provider would be given SSH access, and continue from there. For some providers this serves as a welcome professionalisation of the industry, and a significant improvement from the time (described as only 5-10 years ago in one case) when exploits were frequently untargeted (e.g., to the specific operating system and service patch), launched en-masse, and used without full knowledge of their contents. Other providers were more critical of this shift. Although it was infrequently discussed in stakeholder interviews, for some providers the short length of engagements meant that they felt there was no longer time to do rigorous testing of proof-of-concept exploits before their use in engagements, which could also be a contributor to the move towards scenario-based, no exploitation testing where no "penetration" occurs.

5.2.2. Methodologies

The second concerns methodologies and their limited use by providers, which may reflect on the diversity of service models. As one might expect, no provider followed one particular methodology, and instead considered their methodology to be a synthesis of community standards. Out of the 32 providers, 10 mentioned Open Source Security Testing Methodology Manual (OSSTMM) and 16 mentioned Open Web Application Security Project (OWASP) in general terms, with three providers more specifically mentioning the OWASP Testing Guide. Other methodologies in use during the interviews were the PTES by six providers and those adopted by the National Institute of Standards and Technology (NIST) (again in general terms, without reference to a specific standard) by three providers. One provider noted that NIST 800-115 has gained prominence in the past year, due to the new PCI DSS version 3 standard, but described it as "old and invalid". Providers stated that no public methodologies were followed for high assurance and/or safety-critical environments requiring specialist approaches, such as industrial control systems, as in their experience, none had been created (the ecosystem review of Section 4.3, however, showed a small number exist). It is also worth noting that both the CHECK and CREST schemes require organisations to have defined a methodology that is to be reviewed before acceptance onto the schemes. Although providers mentioned methodologies in this study, in some cases, responses seemed more of an attempt to demonstrate their methodologies received influences from external sources. A selection of elicited responses can be seen below.

> *"We're aligned to OWASP"*
> *"Whatever's available"*
> *"A combination of everybody's"*
> *"Our internal methodology is based on all that*

stuff"

Providers on methodologies

Provider responses did, however, demonstrate that there has yet to be a methodology (for the mainstream target environments: network, web and mobile) that has received widespread adoption within the industry. The lack of a consistent peer reviewed methodology does not necessarily indicate that providers are doing an inadequate job in terms of coverage. Providers frequently stated that they do their best to identify as many vulnerabilities as possible within the duration of the engagement. It does, however, suggest a reluctance to have an external methodology forced upon them. In the opinion of the providers, this allows them the flexibility to tailor their offerings to their clients. It might also be a reflection of the difficulties in defining a methodology in a fast changing, highly complex environment. Despite the lack of utilisation of peer reviewed methodologies, there was the consensus that methodologies play a key role in emerging markets, by improving client education and establishing a rough de facto standardisation of assessment activities. Such methodologies, however, were perceived to best manifest through industry and community efforts.

Standards that specify requirements for the practical assessment can be seen to enforce a form of methodological requirement, and therefore, provider perceptions about this form of standardisation were understandably negative. Perceptions on this topic are best illustrated in light of OWASP ASVS, which as a community developed standard, and a document open to peer review, presents an ideal opportunity to discuss stakeholders views on this form of standardisation. These views can largely be summarised as ignorance or indifference. With respect to the lack of knowledge about OWASP ASVS, multiple providers suggested a lack of awareness of many OWASP projects beyond the Top 10 and Testing Guide. As one provider summarised, "OWASP is interesting. For some reason outside the Top 10 their work is not picked up by the industry". A small number of providers had heard of the OWASP ASVS, and two providers had done tests to it, but only on client requests. For some, ASVS was met with criticism due to its lack of differentiation between white and black box testing, and did not account for the limited willingness for clients to release source code. A counter to this criticism could be made, however, in that it is only the higher levels of verification within the standard that require source code review. A further felt that it would be impossible to meet the requirements of the standard in the short time frames clients are willing to procure for testing, even for their more experienced testers. Other providers were more positive, but never so far as having the desire to use it widely. One provider did state that although they did not use OWASP ASVS, because it was a third party document, it was "good as a sales tool". Another provider felt ASVS's failure to penetrate the market was not so much a failing with the standard itself, but due to a lack of demand in the buying community. "If the buying community is not asking for it, providers won't be willing to spend the time implementing it. If the buying community wanted it, the industry would be all over it."

5.2.3. Social Engineering

Approaches to social engineering engagements were also discussed with providers. Services described fell into two categories: First, scenario-focused in the manner that social engineering is traditionally understood (e.g., with a specific end goal). Second, audit-based social engineering (e.g., to determine the awareness level of a department). For where "human exploitation" occurs within social engineering, providers offering this service described a robust sign-off process, often with multiple stages. A typical provider described process involved scoping and the determination of Open Source Intelligence (OSINT) sources (e.g., social media), the use of OSINT to discover information about an organisation and its employees, the creation of attack scenarios, and the proposal of those scenarios to the client, who would then decide whether they wished the provider to proceed or not. In multiple cases the provider stated that the client would return a list of alternative targets for the scenarios to the ones proposed (e.g., OSINT might reveal information on a C-Suite member, but the client may want to target those in lower positions). Providers were generally adamant about the discussion of any social engineering attack before its use, "otherwise it's easy to get into hot water" and there's a "potentially higher risk for things to go wrong". Most providers described their services as being scenario-focused, with a smaller number being audit-focused, although in some cases offered both.

Client sign-off opens a separate issue; the ethics and legal aspects of social engineering. Two providers described situations where it was the client asking the provider for testing with potentially ethically dubious motivates (e.g., perform a phishing attack on a department, crack the passwords, and provide the names of the 10 individuals with the worst passwords). These providers stated that it is often them as testers that have to inform the client that such behaviour would not be ethical. Despite this, the majority of providers did state that there was complex negotiations before social engineering occurs, which often involves a client organisation's Human Resources (HR) department, or at the minimum, the provider suggests the client contact their HR department before testing. Anonymisation of victims of social engineering received strong responses from both perspectives. Of the providers that gave a clear answer, five said no (across two organisations) and seven said yes (across six organisations) to anonymising results. One provider that said no, stated: "It would be wrong to anonymise anything. The client commissioned the test, and the client owns the results when you give it to them". The argument for anonymisation was largely that it is a training and policy failing and "not a finger pointing exercise". In one case, the provider stated that they request

HR be at debrief meetings to emphasise this point. Sometimes anonymisation is inadequate as it is easy to determine the target (e.g., the receptionist at the front desk on a particular day and time). In these situations some of the "yes" providers stated that they try to obfuscate results. One provider argued that due to the ease in which social engineering attacks succeed, they do agree that one person within the client organisation can know the names, to allow for technical remediation (e.g., resetting passwords). A further provider stated that they get the client to sign a document stating that they will not take action against employees found to be targets of the test. Other providers could not give a hard yes or no answer to anonymisation, but general practice is that they "tend not to give names". Two of the providers did state they have felt pressure from client organisations to disclose names. In one case they stated that sometimes the client asks informally and they "may tell them, may not".

With respect to the legal and ethical aspects when conducting social engineering assessments, one provider summarised this as "a minefield". This is notably the case in scenario-driven tests that involve any form of physical penetration. There was a strong consensus amongst stakeholders that the provider community would benefit from a synthesis of ethical and legal material on this topic into a common source, such as a set of guidelines or a code of ethics. Providers views included "it might be useful","if there was one it would be really useful","some framework at the level of ethics would be useful" and"it needs to be done - I can imagine the arguments though". One provider did state they had previously searched for guidelines but without success. Another argued that "a code of ethics in the UK would likely see some success and approval" due to the requirement for being "whiter than white" as a penetration tester, compared to other countries in the world where a criminal history is sometimes encouraged or overlooked. Clients also expressed a desire to see greater ethical and legal guidelines. One client argued that "at that level they're not engaging with customers but its employees on an interpersonal level", with another client adding, "there should be guidelines. For the protection of those doing the testing as much as anything else". CREST was the body of choice for any guidance for two providers, with the suggestion that it integrates with their complaints process. "[It] should definitely be from CREST". Some providers, including those who were pro-guidelines, did express that such an endeavour must be cautious not to constrict the industry and the value of testing. During one client interview, the interviewee "could not comment" on whether social engineering had been conducted in tests on their organisation, although could discuss other technical tests (e.g., network and web application penetration testing). One could argue that this is perhaps indicative of how people perceive social engineering and the sensitivity of human-focused testing, which strengthens the need for clarification on ethical and legal guidelines.

5.3. Post-Engagement

> *"I've never seen any wow reports, but a lot of bad ones"*
> *"Shocking"*
> *"Generally very hit and miss"*
> *"Appalling"*
> **Providers on the reports of other providers**

"Underwhelming" was the overarching theme in the perceptions of reporting from providers and clients. Providers expressed satisfaction with the quality of their own reports, but had largely disapproving opinions of the reports produced by other providers. A small number of providers felt there was some consistency between their direct competitors, with one large provider arguing that a level of consistency had been achieved through the movements of individuals between provider organisations.

> *"The quality varies immensely ... the quality can be atrocious"*
> *"Often basically a Nessus output in PDF format"*
> *"Very impressed"*
> *"... great deal of variability"*
> *"Some are atrocious; others well thought out"*
> *"The quality of the document was high"*
> *"No significant quality variation"*
> *"Some are so shocking, it's hilarious"*
> **Clients on reporting quality**

Client interviews highlighted a significant perceived variability in the quality of reports from providers. The above quotes were extracted from the views of eight clients in eight organisations. One interesting finding was that the smaller clients had the best opinions on the quality of reporting. Generally, the larger the client (typically, therefore, with a greater in-house IT capability), the greater the perceived variability.

Two widely cited issues that will not be discussed in detail here are the mis-marketing of vulnerability assessment services as penetration testing services, and the "quality" of report content. The former is a systemic issue that stems from pre-engagement negotiations. The latter is about individual capability, which stakeholders strongly felt should be the responsibility of technical bodies.

5.3.1. Reporting Structure

All providers were found to follow a similar high-level reporting structure. At its most basic, all reports were described to have a managerial and technical section. Managerial sections typically contained the executive summary and engagement details (e.g. scope). Clients were moderately satisfied with provider efforts, but many felt managerial sections were still too technical, and often needed rewriting for internal communications. The technical section broadly contained the lists of discovered vulnerabilities and recommendations. The provider's implementations for both were varied. Some best practices for reporting structure that were noted include the use of document histories, information on providers involved in testing (e.g. qualifications), attack narratives (e.g., a descriptive explanation of what route was taken by the assessor in attacking the target), root cause recommendations and appendices of test data (e.g. logs of tool outputs and systems "touched" during testing).

5.3.2. Vulnerability Scoring

The first major issue highlighted was the diverse use of default metrics for scoring vulnerabilities. The Common Vulnerability Scoring System (CVSS) version 2.0 was mentioned frequently, and was often mandated by some clients; however, providers were critical of it in its current form, arguing that it was only suitable for certain technologies, its scores often did not reflect real-world risks, and that it failed to account for vulnerability chains (e.g. multiple medium risk vulnerabilities created one of high risk) or the presence of mitigating security controls. Instead, providers frequently described the use of alternative metrics, such as: qualitative scores (usually high, medium and low); impact to Confidentiality, Integrity or Availability (CIA); ease of exploitation; proprietary CVSS derivatives; or a combination of multiple metrics in a matrix. For clients, the variety of scoring mechanisms was found to be problematic for tracking performance over time and comparing results between providers. Furthermore, issues were felt to be compounded due to the subjectivity in arriving at a particular score, such as when providers tried to adapt CVSS to account for its aforementioned limitations, or address one or more aspects using their own metric system. The survey highlighted a strong opposition to the potential for mandating a specific metrics system, as this is where providers felt their value was generated; however, some providers did feel that clients mandating the inclusion of unmodified CVSS scores regardless of the use of another "main" metric system would provide a "quick win" for consistency within the industry. Version 3.0 of CVSS is currently in development, and some providers expressed the hope that this would lead to a natural resolution and improvement of this issue.

5.3.3. Recommendations

Another area of concern was the quality and content of recommendations. Smaller clients were the most satisfied, with larger clients having more qualms. Frequent criticisms included the lack of prioritisation (beyond the implicit prioritisation based on the vulnerability score), categorisation, and root cause analysis. Root cause analysis featured heavily in client demands, but providers were largely seen to be failing to deliver on this. One client was particularly critical of CHECK reports for their lack of root cause analysis, stating that it "rarely happens in their CHECK reports" and that there is "no interpretation of results". Only seven providers (six organisations) stated that they included any root cause analysis in any form of penetration test report, although more did state that recommendations were prioritised. One of the largest clients in the project went further to argue that providers need to include scenarios in their root cause analysis to enable a greater understanding of vulnerability chains and their impact. Interestingly, a criticism of clients that arose at multiple points within the study was the claim that they often only spot fix, rather than address root causes, and that issues continue to appear in subsequent engagements (e.g. their yearly audit). While the ultimate responsibility to address systemic issues lies with the client, based on the findings in this study, it would be difficult to claim that many providers are going to great lengths to facilitate this.

5.3.4. Validation of Vulnerabilities

The final issue was once a client has implemented remediation to vulnerabilities, they then have two options for validating its efficacy. They could either obtain a retest of the vulnerabilities (usually at an additional cost) or test for the vulnerability themselves. Most clients of penetration testing engagements do not have the skills or training to understand and recreate the vulnerabilities themselves, which therefore means they must be empowered to do so. Only nine providers (seven organisations) stated that proof-of-concepts were included within their reports (e.g. a single command or script that can be queried or executed to demonstrate the issue), with clients describing their presence as rare. The majority of providers offered retests instead, although some providers stated that some information was provided, such as "what tools were used". The majority of clients expressed an interest in proof-of-concepts being made available, with some clients stating that they would also like to see attack narratives. One client stated that this was because remediation is often undone, and attack narratives would facilitate better understanding of cyber threats and empower them to implement more effective mitigating security controls. A provider argued that not including narratives or proof-of-concepts in reports was a "business decision" by provider organisations, with another suggesting the same issues, arguing that this information will typically be produced to enable them to conduct a retest anyway. The provision of such information aids in educating the client to improve their security, but doing so would not be financially beneficial for providers.

5.3.5. Improving Reporting

The difficulty in achieving greater quality of reporting is balancing the need for consistency with the resistance to standardisation and the providers' desire to maintain flexibility in the reporting process. Auditing and setting guidelines were two methods suggested by providers that could help to achieve this.

CHECK reports are reviewed by CESG for quality and metrics. Any issues found are raised with the customer and/or the CHECK company. This is supplemented with an annual audit where CHECK companies are requested to send two examples of work that best demonstrate their technical ability. CREST reports are not audited. Two providers (one CREST organisation) argued that "they should be doing them." However, without regulatory or other external support (e.g. as with the CHECK scheme) such an approach could see opposition from providers. In part, because a shift in the governance framework may require adaptations to business models, but also due to the practical challenges of implementing this in the private sector (e.g. handling client confidentiality).

Authoritative guidelines on reporting best practices were suggested in the belief that if clients had access to such guidelines, their expectations would raise the reporting standards by providers. PTES does contain reporting standards; however, PTES has failed so far to achieve widespread awareness amongst the buying community. The provider community is aware of issues around reporting, and some providers are taking steps to address this. One example that the authors were made aware of during the study was a community project that aims to create a baseline, minimum standard for reporting. The output will be a series of guidelines outlining best practices, and an example report that will be made available to the public (i.e. both providers and clients). The example report will be produced based on the findings of a real engagement undertaken by the project's group. This project involves some providers within the PTES group; however, this project will be independent from PTES.

5.4. Summary of key findings

Several key findings on the different stages of the engagement process warrant highlighting:

- Pre-engagement:
  - The ambiguity of what constitutes a *penetration test* was a cause of confusion and frustration amongst stakeholders.
  - Smaller providers utilise face-to-face meetings to distinguish themselves against the "faceless" larger providers.
  - Undisclosed third party systems were a common cause of delayed testing.
  - More effort needs to be made to provide guidance for international engagements.
- Practical assessment:
  - Many providers and clients have an aversion to exploitation of vulnerabilities found, with the short length of engagements a contributing factor.
  - There was a diverse, and limited, use of existing methodologies by providers, and a reluctance to have an external methodology forced upon them.
  - Both scenario-focused and audit-based social engineering were used in engagements, with scenario-focused being most prevalent.
  - The ethics around the use of social engineering was a grey area, with a range of opinions on anonymisation, and the providers often having to inform the clients that a request would amount to unethical behaviour. There was a strong call for more guidance in this area.
- Post-engagement:
  - Most reports followed a similar high-level structure.
  - There was a diverse use of different metrics for scoring vulnerabilities, and a level of subjectivity, which made it difficult for clients to track performance over time and compare results between providers.
  - There was a strong objection by providers to mandating a specific metrics system.
  - There was a lack of prioritisation, categorisation, and root cause analysis of recommendations, which made it difficult for clients to understand the impact, and address systemic issues.
  - Clients want to see proof-of-concepts of vulnerabilities and attack narratives, but these are rarely provided in reports.
  - Auditing and setting guidelines were suggested as routes to improving the quality of reports.

6. Opportunities for Ecosystem Advancement

This analysis has identified a multitude of opportunities for further discussion, analysis and improvement within the simulated security assessment ecosystem, most notably at the beginning and end of engagements. Based on these findings seven development areas are proposed: three which involve standardisation, and four which the industry (and wider security community) itself is in the most suitable position to evoke change. Given the importance (and rapid growth) of simulated security assessments, resolving these needs for best practice quickly would aid providers and buyers. The seven development areas can be seen in Figure 5, and are detailed within the following sections. Each development area is numbered according to the relevant engagement phase: pre-engagement (P1), practical assessment (P2), and post-engagement (P3). Certain development areas provide contributions towards addressing the requirements of other areas; contributing relationships and their directions are also illustrated within Figure 5 using arrowed lines.

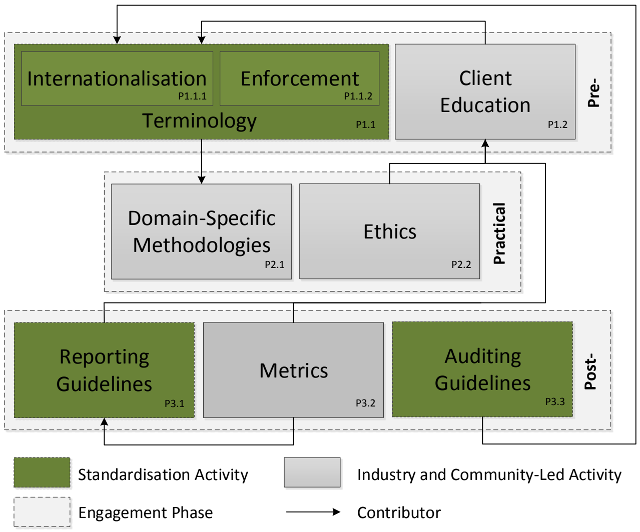

Figure 5: Opportunities for Improvement within the Simulated Security Assessment Ecosystem

6.1. Pre-Engagement

Providers offer a diverse mix of qualities and depth of simulated security assessments. The buying market in the words of one provider is in need of something to compare "like-for-like". Two development areas are proposed to address this requirement.

Standardisation of terminology (P1.1) is recommended to enable clients to make more informed procurement decisions. The current heterogeneity of service offerings creates uncertainty for clients in what will be delivered, which establishes high requirements for pre-procurement due diligence, which is not always feasible. The form of standardisation proposed here is not to attempt to establish wholesale consistency between providers, which would be both infeasible to achieve and may lead to a form of commoditisation on its own. Instead, a form of standardisation is proposed to establish and hold providers to a minimum service quality for the different service definitions (e.g., vulnerability assessment, penetration test, and red team exercises), while allowing the flexibility to customise services to meet client requirements. The PTES is one notable community project, which is working towards a similar goal of terminology standardisation. Standards bodies should look towards developing relationships with community efforts to achieve similar terminology models. Such an approach leverages existing work by subject matter experts. Furthermore, in the opinions of providers, the reputation of standards bodies can aid in bringing these concepts to the mass market.

Internationalisation (P1.1.1) Stakeholders argued that a focus on standardisation to delineate service definitions would aid in addressing the commoditisation of simulated security assessments, whilst not being tied to a specific region, and thus open to internationalisation. From the perspective of providers, the role of terminology standardisation in the internationalisation of simulated security assessment services was significant. Although such a standard would be intended to have a positive impact within the UK market, it would have wider impact within the fledgling international markets that providers are increasingly looking to for growth.

Enforcement (P1.1.2) Standards are ineffective without enforcement. The manner in which such standardisation is enforced requires further research and discussion. Stakeholder opinions within this study followed two tracks. First, having existing technical bodies place greater emphasis on policing service quality. However, two challenges would need to be resolved. The first is the challenges of practically conducting such policing; some of which were discussed in the context of reporting in Section 5.3. The second concerns internationalisation, and how these standards are tied to largely region specific certifications (e.g., CHECK is UK only, while the two implementations of CREST, CREST UK and CREST Australia, operate independently, with expansion also in progress to Singapore). The potential for an international standard in the traditional sense, which is enforced by national accreditation bodies through certification bodies requires further research (in this scenario existing technical bodies would be certification bodies who then certify provider companies).

Client education (P1.2) Standardisation of terminology alone, however, is not a panacea to pre-engagement woes. Client engagement is paramount. A number of development areas contribute to this aim, but here the call from stakeholders was for providers to place greater emphasis on not only educating clients on the importance of security, but empowering them through education to improve their security posture. Such empowerment occurs throughout the engagement life cycle: it begins with education on effective and appropriate service models, and ends with education on remediation. For pre-engagement activities, such education is largely self-completing. Service understanding develops naturally through exposure to services (e.g., annual assessments). However, such development can be facilitated by providers through greater transparency within their service models.

6.2. Practical Assessment

A strong opposition to any form of standardisation in relation to the practical assessment phase has been discussed, along with the success, and continued improvement of the technical bodies (e.g., CHECK and CREST) in this domain. Despite this, two areas for potential improvement are noteworthy.

Domain-specific methodologies (P2.1) Providers scarcely use peer-reviewed methodologies as described in Section 5.2.2, instead preferring a synthesis of approaches for their internal methodology. This synthesis is significant, as despite their lack of strict usage, public methodologies do guide the design of internal methodologies. In most scenarios, such methodologies are well established (e.g., infrastructure, web applications, and increasingly, mobile devices). For engagements involving niche and novel environments, however, this is atypical. Furthermore, in such engagements providers expressed a notable malleability in the service definitions used by some providers, which it can be argued, may arise through the lack of peer-reviewed knowledge on what assessments should entail, along with establishing further evidence on the requirement for standardisation of terminology (P1.1). The development of domain-specific methodologies is also recommended to aid in addressing this knowledge gap. One notable case in which this occurs is with Industrial Control Systems. Since the conclusion of this paper's interview process, CREST has begun undertaking a project to this effect. This project is examining the need for new and updated public methodologies, along with exploring the possibility and form of services being expanded into this domain (e.g., the STAR scheme as a service and as individual qualifications).

Ethics (P2.2) One success of the industry within the UK (notably by CREST) has been the formalised structure for dealing with complaints at the technical body level. Included within this structure is the facility to handle complaints concerning ethics. However, for this system to function effectively, providers must have a clear understanding of the ethical framework within which to work under. In broad terms, providers felt the industry has proven capable of handling this responsibility in the majority of cases. However, a need for further research around *ethics* (P2.2) was identified in multiple areas; most notably for where simulated security assessments involve human subjects (e.g., where social engineering is used). In this scenario, research would also be intended to improve client education (P1.2) on their responsibilities to their employees (e.g., with respect to anonymisation, and remediation being training-led rather than punitive). The topic of ethics has been addressed somewhat within academic literature (see Section 2.4), e.g. for phishing experiments. Further research is needed, and for the researched methodologies to be taken up by industry.

6.3. Post-Engagement

Post-engagement comes in a variety of forms. It is both the immediate aftermath of an engagement, and the implications of that engagement. Here three development areas are proposed that span both categories.

Although the high-level structure of reports was found to be relatively standardised between providers, at a lower level, stakeholders were found to be dissatisfied with various characteristics of reports. For providers, it was predominantly their experiences of seeing competitors offering vulnerability assessments that had been mis-marketed and sold as penetration tests. The most effective resolution to this arguably comes not from a focus on the report itself, but around the standardisation of terminology (P1.1). For clients, it was the inconsistency between providers, and a lack of depth in the provision of metrics and recommendations (e.g. root cause analysis) to empower the client to understand more about the security issues within their environment.

Reporting guidelines (P3.1) A standard in the form of a guideline, rather than a specification, for reporting is recommended. Rigid standardisation would likely see significant opposition; providers see their reports as a means to differentiate themselves and add value to their offering to the client. However, guidelines describing best practices could be produced, and have the potential to provide an effective alternative. Through standards bodies, such guidelines would likely gain significant exposure within the buying community, which may be otherwise difficult to achieve. Exposure facilitates *client education* (P1.2), which empowers clients to make informed decisions when interacting with providers, and gives them a clear conception of what they should expect. Such guidelines could address all of the aforementioned issues within this study, while describing best practices around the processes that support report production (e.g. quality assurance). As with the recommendations on terminology, standards bodies should look to leverage existing efforts within the community to raise reporting standards. This should include working with technical bodies in the UK, such as CREST, while remaining aware that guidelines should not be region-specific.

Metrics (P3.2) Of the characteristics of reports that elicited dissatisfaction, one of the most prominent was that of security metrics. To some extent the issues surrounding metrics can be mitigated through reporting guidelines (P3.1) and the wider educational process for clients (P1.2). If clients are educated and empowered to mandate metrics and related reporting requirements, as many larger enterprises have done, a degree of consistency can be achieved. Providers, however, often perceived such an approach negatively (e.g., as these metrics differ highly between clients), and furthermore, the general consensus was that providers see unique metric approaches as a market differentiator. It is through the metrics after all that security is measured and understood. For clients, however, the call was for that understanding to be facilitated by consistent measurements between providers. Two factors require further research and discussion. The first is establishing consensus and backing for a consistent measurement approach. A large part of the debate arises through providers perceiving that there is no "good" metric. The notion of mandatory CVSS 2.0 (supplementary to any other approach), and hope for CVSS 3.0 was touted by a number of providers. The second is to examine how this can be achieved while minimising the subjectivity involved in making such judgements, which is a primary source of inconsistency.

Auditing guidelines (P3.3) Looking beyond the scope of an engagement, simulated security assessments contribute to ISO/IEC 27001 audits under an isolated ISO/IEC 27002 security control "technical compliance review". Auditing guidelines have been produced previously in ISO/IEC 27008; however, they are being revised. Standards bodies should look to provide auditing guideline that establish a clearer link between the scope of an engagement and its findings, and ISO/IEC 27002 security controls beyond the narrow categorisation of a technical compliance review. Some stakeholders mentioned the perception of ISO/IEC 27001 being a "check box exercise", and where simulated security assessments are of relevance in the audit process, the audit is merely a confirmation that it has occurred, rather than a detailed analysis of its findings to determine whether the security controls that have been implemented are consistent with the objectives of the ISMS.

Auditing guidelines provide the opportunity to link the socio-technical security controls of ISO/IEC 27002, and the socio-technical nature of simulated security assessments. Such assessments may only be one group of assessment methodologies, but they continue to rise in popularity and are increasingly seen as a regulatory requirement. Furthermore, penetration tests and red team exercises are arguably the most realistic methodologies currently available for simulating cyber threats. As part of this process, it is recommended that standards bodies examine the integration of simulated security assessments with other standards, such as ISO 31000. A diverse array of metrics can be used as part of an engagement, but what is its meaning for risk management, and how does this impact the risk that is to be managed as part of ISO/IEC 27001? Furthermore, auditing guidelines maintain a close relationship with the requirement for terminology standardisation (P1.1). If a standard or other assurance scheme mandates a particular variety of simulated security assessment, how can auditors ensure that it has been appropriately delivered without consensus establishing what such an assessment should look like?

7. Conclusion

The CHECK and CREST schemes, along with technical bodies such as the Tigerscheme, have successfully defined the technical capabilities of individuals who perform simulated security assessments, and can be seen to be making great efforts to encourage evolution within the industry. In addition, both CHECK and CREST have laid the foundations for the assessment of organisational processes that support engagements. The professionalisation of simulated security assessments that such schemes have enabled is primarily concerned with the UK market. The findings of this study suggest that on an international scale, the level of professionalisation is less formalised, with respect to both individuals and provider organisations (e.g., in terms of how evidence of competency can be provided to employers and clients), although there are isolated exceptions that were highly regarded, such as is the case with some individual qualifications within the United States (specifically those from Offensive Security). It can therefore be argued that the international market can learn many lessons from the path to professionalisation that has been paved by the UK market. There is much evidence to suggest that a shift towards a UK-style professionalisation is not only possible, but desired, as can be seen through the recommendation of UK originating schemes in non-UK and international standards (e.g., PCI DSS [46]), and as is evidenced by the expansion of UK originating schemes to other countries (e.g., CREST Australia).

Despite the professionalisation, this study has identified that there remains a number of issues at the start and end of the engagement process that the industry has currently failed to address. This is not through a lack of awareness of these issues; this study has highlighted that both providers and clients are dissatisfied by the lack of transparency and consistency in industry offerings. It is based on these findings that the authors make the proposal: *standardisation is needed.*

Standards must be well formed to avoid the potential to suffocate and hinder rapidly evolving industries, such as the one we find with simulated security assessments. Self-regulation in such an environment is an ideal solution (e.g., through trade organisation's auditing services delivered as opposed to static document reviews of methodologies); however, as one provider stated, when it comes to current industry offerings, it can be a "Wild West". This is not due to a lack of technical capability within the industry, as the technological bar has been set and maintained by the technical bodies. One provider argued that the "UK security industry can provide anything the market asks for"; the problem is that "it [the market] does not ask the right questions", which promotes ambiguity when interpreting service model requirements and results in a lack of demand for robust governance structures. The framework for future improvement proposed here is intended to work towards remediating industry issues, using a collaborative approach of standardisation and industry-led development.

The form of standardisation proposed within these recommendations must be shaped through continued discussion with all stakeholders. As such, the findings of this paper have led to a preliminary workshop in July 2015 which was hosted by BSI to determine the consensus between stakeholders in the UK. Despite the vocal findings on the need for standards from many stakeholders the feedback from the workshop was that the two key stakeholder groups for the UK (CESG and CREST) would need to be explicitly on board for this to proceed as a national standard. The potential for ISO/IEC standardisation is also being explored.

Acknowledgements

The authors would like to thank the EPSRC through its financial support for Lancaster University's Impact Acceleration Account (Grant Reference EP/K50421X/1).

References

[1] J. Such, A. Gouglidis, W. Knowles, G. Misra, A. Rashid, Information Assurance Techniques: Perceived Cost Effective-

ness, Elsevier Computers and Security 60 (2016) 117–133. doi:doi:10.1016/j.cose.2016.03.009.

[2] H. J. De Vries, Standardization: A business approach to the role of national standardization organizations, Springer Science & Business Media, 2013.

[3] P. Swann, The economics of standardization (2000).

[4] P. Swann, The economics of standardization: an update (2010).

[5] K. Blind, S. Gauch, R. Hawkins, How stakeholders view the impacts of international ICT standards, Telecommunications Policy 34 (3) (2010) 162–174.

[6] K. Blind, The impact of standardization and standards on innovation (2013).

[7] W. Knowles, A. Baron, T. McGarr, Analysis and recommendations for standardization in penetration testing and vulnerability assessment: Penetration testing market survey, Tech. rep., British Standards Institution (BSI) (2015).
URL http://shop.bsigroup.com/forms/PenTestStandardsReport/

[8] G. Mcgraw, Software security, IEEE Security & Privacy Magazine 2 (2) (2004) 80–83. doi:10.1109/MSECP.2004.1281254.

[9] G. McGraw, Software Security, Datenschutz und Datensicherheit - DuD 36 (9) (2012) 662–665.

[10] K. van Wyk, G. McGraw, Bridging the Gap between Software Development and Information Security, IEEE Security and Privacy Magazine 3 (5) (2005) 75–79. doi:10.1109/MSP.2005.118.

[11] J. Such, A. Gouglidis, W. Knowles, G. Misra, A. Rashid, The Economics of Assurance Activities, Tech. Rep. SCC-2015-03, Security Lancaster, Lancaster University (2015).

[12] B. Arkin, S. Stender, G. McGraw, Software penetration testing, IEEE Security and Privacy Magazine 3 (1) (2005) 84–87. doi:10.1109/MSP.2005.23.

[13] G. Hardy, The relevance of penetration testing to corporate network security, Information Security Technical Report 2 (3) (1997) 80–86. doi:10.1016/S1363-4127(97)89713-0.

[14] M. Bishop, About Penetration Testing, IEEE Security & Privacy Magazine 5 (6) (2007) 84–87. doi:10.1109/MSP.2007.159.

[15] D. Geer, J. Harthorne, Penetration testing: a duet, in: 18th Annual Computer Security Applications Conference, 2002. Proceedings., IEEE Comput. Soc, 2002, pp. 185–195. doi:10.1109/CSAC.2002.1176290.

[16] P. Midian, Perspectives on Penetration Testing, Computer Fraud & Security 2002 (6) (2002) 15–17. doi:10.1016/S1361-3723(02)00612-7.

[17] C. Herley, W. Pieters, "if you were attacked, you'd be sorry": Counterfactuals as security arguments, in: Proceedings of the 2015 New Security Paradigms Workshop, NSPW '15, ACM, New York, NY, USA, 2015, pp. 112–123. doi:10.1145/2841113.2841122.
URL http://doi.acm.org/10.1145/2841113.2841122

[18] A. Tang, A guide to penetration testing, Network Security 2014 (8) (2014) 8–11. doi:10.1016/S1353-4858(14)70079-0.

[19] J. Yeo, Using penetration testing to enhance your company's security, Computer Fraud & Security 2013 (4) (2013) 17–20. doi:10.1016/S1361-3723(13)70039-3.

[20] K. Xynos, I. Sutherland, H. Read, E. Everitt, A. Blyth, Penetration Testing and Vulnerability Assessments: A Professional Approach, in: International Cyber Resilience Conference, 2010, pp. 126–132.
URL http://ro.ecu.edu.au/icr/16

[21] C. P. Pfleeger, S. L. Pfleeger, M. F. Theofanos, A methodology for penetration testing, Computers & Security 8 (7) (1989) 613–620. doi:10.1016/0167-4048(89)90054-0.

[22] A. Bechtsoudis, N. Sklavos, Aiming at Higher Network Security through Extensive Penetration Tests, IEEE Latin America Transactions 10 (3) (2012) 1752–1756. doi:10.1109/TLA.2012.6222581.

[23] J. N. Goel, B. Mehtre, Vulnerability Assessment & Penetration Testing as a Cyber Defence Technology, Procedia Computer Science 57 (2015) 710–715. doi:10.1016/j.procs.2015.07.458.

[24] J. P. McDermott, Attack net penetration testing, in: Proceedings of the 2000 workshop on New security paradigms - NSPW '00, ACM Press, New York, New York, USA, 2000, pp. 15–21. doi:10.1145/366173.366183.

[25] H. Thompson, Application penetration testing, IEEE Security and Privacy Magazine 3 (1) (2005) 66–69. doi:10.1109/MSP.2005.3.

[26] T. Dimkov, A. van Cleeff, W. Pieters, P. Hartel, Two methodologies for physical penetration testing using social engineering, in: Proceedings of the 26th Annual Computer Security Applications Conference, ACSAC '10, ACM, New York, NY, USA, 2010, pp. 399–408.

[27] L. Guard, M. Crossland, M. Paprzycki, J. Thomas, Developing an empirical study of how qualified subjects might be selected for IT system security penetration testing, Annales UMCS Sectio AI Informatica 2 (1) (2015) 414–424.

[28] T. Caldwell, Ethical hackers: putting on the white hat, Network Security 2011 (7) (2011) 10–13. doi:10.1016/S1353-4858(11)70075-7.

[29] P. Y. Logan, A. Clarkson, Teaching students to hack: curriculum issues in information security, in: Proceedings of the 36th SIGCSE technical symposium on Computer science education - SIGCSE '05, ACM Press, New York, New York, USA, 2005, pp. 157–161. doi:10.1145/1047344.1047405.

[30] S. A. Saleem, Ethical hacking as a risk management technique, in: Proceedings of the 3rd annual conference on Information security curriculum development - InfoSecCD '06, ACM Press, New York, New York, USA, 2006, pp. 201–203. doi:10.1145/1231047.1231089.

[31] B. A. Pashel, Teaching students to hack: ethical implications in teaching students to hack at the university level, in: Proceedings of the 3rd annual conference on Information security curriculum development - InfoSecCD '06, ACM Press, New York, New York, USA, 2006, pp. 197–200. doi:10.1145/1231047.1231088.

[32] D. Jamil, M. N. A. Khan, Is Ethical Hacking Ethical?, International Journal of Engineering Science and Technology 3 (5) (2011) 3758–3763.

[33] B. Smith, W. Yurcik, D. Doss, Ethical hacking: the security justification redux, in: IEEE 2002 International Symposium on Technology and Society (ISTAS'02). Social Implications of Information and Communication Technology. Proceedings (Cat. No.02CH37293), IEEE, 2002, pp. 374–379. doi:10.1109/ISTAS.2002.1013840.

[34] A. Matwyshyn, A. Keromytis, S. Stolfo, Ethics in security vulnerability research, IEEE Security & Privacy Magazine 8 (2) (2010) 67–72.

[35] J. Pierce, A. Jones, M. Warren, Penetration Testing Professional Ethics: a conceptual model and taxonomy, Australasian Journal of Information Systems 13 (2) (2006) 193–200. doi:10.3127/ajis.v13i2.52.

[36] Q. M. Ashraf, M. H. Habaebi, Towards Islamic Ethics in Professional Penetration Testing, Revelation and Science 3 (2) (2013) 30–38.

[37] F. Mouton, M. M. Malan, K. K. Kimppa, H. Venter, Necessity for ethics in social engineering research, Computers & Security 55 (2015) 114–127. doi:10.1016/j.cose.2015.09.001.

[38] P. Finn, M. Jakobsson, Designing ethical phishing experiments, Technology and Society Magazine, IEEE 26 (1) (2007) 46–58.

[39] R. Reece, B. Stahl, The professionalisation of information security: Perspectives of UK practitioners, Computers & Security 48 (2015) 182–195. doi:10.1016/j.cose.2014.10.007.

[40] Department for Business Innovation and Skills, Cyber Essentials Scheme: Summary, Tech. rep. (2014).
URL https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/317480/Cyber_Essentials_Summary.pdf

[41] PCI Security Standards Council, Payment Card Industry (PCI) Data Security Standard: Requirements and Security Assessment Procedures (Version 3.0), Tech. rep. (2013).
URL https://www.pcisecuritystandards.org/documents/PCI_DSS_v3.pdf

[42] British Standards Institution, BS ISO/IEC 15408-1:2009. Information technology - Security techniques - Evaluation criteria for

IT Security (2014).

[43] British Standards Institution, BS ISO/IEC 15408-3:2008. Information technology - Security techniques - Evaluation criteria for IT security. Part 3: Security assurance components (2009).

[44] CREST, A Guide to the Cyber Essentials Scheme, Tech. rep. (2014).
URL http://www.crest-approved.org/wp-content/uploads/Crest-Cyber-Essentials-Guide-final.pdf

[45] CESG, Cyber Essentials PLUS: Common Test Specification V1.2, Tech. rep. (2014).
URL https://www.cesg.gov.uk/publications/Documents/cyber-essentials-test-specs.pdf

[46] PCI Security Standards Council Penetration Test Guidance Special Interest Group, Information Supplement: Penetration Testing Guidance, Tech. rep. (2015).
URL https://www.pcisecuritystandards.org/documents/Penetration_Testing_Guidance_March_2015.pdf

[47] CREST, CBEST Implementation Guide, Tech. rep. (2014).
URL http://www.bankofengland.co.uk/financialstability/fsc/Documents/cbestimplementationguide.pdf

[48] CREST, An introduction to CBEST, Tech. rep. (2014).
URL http://www.bankofengland.co.uk/financialstability/fsc/Documents/anintroductiontocbest.pdf

[49] NIST, Special Publication 800-115. Technical Guide to Information Security Testing and Assessment, Tech. rep. (2008).

[50] P. Herzog, OSSTMM 3. The Open Source Security Testing Methdology Manual: Contemporary Security Testing and Analysis., Tech. rep., ISECOM (2014).
URL https://scadahacker.com/library/Documents/Assessment_Guidance/OSSTMM-3.0.pdf

[51] U.S. Department of Homeland Security, Cyber Security Assessments of Industrial Control Systems, Tech. Rep. April (2011).

[52] J. Searle, NESCOR Guide to Penetration Testing for Electric Utilities, Tech. rep.
URL http://smartgrid.epri.com/doc/NESCORGuidetoPenetrationTestingforElectricUtilities-v3-Final.pdf

[53] InGuardians, Advanced Metering Infrastructure Attack Methodology, Tech. rep. (2009).
URL http://inguardians.com/pubs/AMI_Attack_Methodology.pdf