Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025
Paper ID #48344

A Standardized Methodology for Evaluating a Digital Badging System

Lt. Benjamin Pederson, Air Force Institute of Technology

Benjamin Pederson is a Cyber Officer in the United States Space Force. He received a M.S. in Cyber Operations from the Air Force Institute of Technology in 2025, and a B.S. in Computer Science from the United States Air Force Academy in 2023.

Dr. Mark Reith, Air Force Institute of Technology

Mark G. Reith is an Assistant Professor of Computer Science at the Air Force Institute of Technology (AFIT). He received a B.S. degree in Computer Science from the University of Portland, Portland, OR, USA, in 1999, a M.S. degree in Computer Science from the AFIT, Wright-Patterson Air Force Base, OH, USA, in 2003, and a Ph.D. degree in Computer Science from the University of Texas at San Antonio, San Antonio, TX, USA, in 2009.

Dr. David Long, Air Force Institute of Technology

David S. Long is an Assistant Professor of Systems Engineering at the Air Force Institute of Technology and a Principal Systems Engineer at KBR. He previously served 25 years in the US Air Force in a variety of positions. His BS Industrial Engineering and Management (North Dakota State University), MS Engineering (California State University, Northridge), and PhD Engineering Systems (Massachusetts Institute of Technology).

Ralucca A. Gera, Naval Postgraduate School

Ralucca Gera is a Professor of Mathematics and the Chair of Applied Mathematics Department at the Naval Postgraduate School. She is a Distinguished graduate of Western Michigan University, MI, USA, where she earned her B.S., M.A. and Ph.D. in Mathematics in 2000, 2002, and 2005 respectively.

Edward D White, Air Force Institute of Technology

Edward D. White is a Professor of Statistics in the Department of Mathematics and Statistics at AFIT. He received his BS in Mathematics from the University of Tampa, MAS from The Ohio State University, and PhD in Statistics from Texas A&M University. His primary research interests include statistical modeling, simulation, and data analytics. E-mail address: edward.white@us.af.mil

Jonathan Zemmer, Air Force Institute of Technology

Jonathan Zemmer is the Educational Technology Liaison at the Air Force Institute of Technology (AFIT). He received a B.A. in Multimedia Technology from Cedarville University, an M.Ed. in Curriculum and Instruction from the University of Cincinnati, and an Ed.D. in Organizational Studies from Wright State University in 2024. His research interests lie in the learning sciences, with a focus on the Social Theory of Learning and educational technology in graduate and military education contexts.

A Standardized Methodology for Evaluating a Digital Badging System

Abstract: Digital badges, a form of micro-credentials, have grown in popularity over the past decade. However, few standard processes exist to assess the potential of digital badging systems within an organization. This study proposes a generalizable methodology for comparing a badging system with other methods of recording skills and competencies. The experimental design is tested using the military's cyber operations community as the target organization. Finally, mixed-method data from thirty-six participants is analyzed in accordance with the methodology. Based on the results, digital badging systems are perceived to be more valuable and usable than a current method of military talent management. This approach supports efforts to enhance formal and informal learning, competency-based learning, granular decision-making, and to build trustworthy systems of record.

Keywords: Digital Badges, Competency-Based Learning, Methodology, Talent Management

Background

Academic institutions and military organizations have expressed interest in digital credentials as a means of recognizing achievement and competency (Chan, 2023; Patel, 2023; L. Facey-Shaw et al., 2018; Education and Training Command, 2023). However, few studies formally evaluate the value associated with this technology. The current research lacks a standardized process for assessing the potential of digital badging within an organization. This research outlines a novel methodology, based upon a systematic review of the literature, to evaluate a digital badging system aimed at improving talent management outcomes within the military learning community. This methodology describes a repeatable process for comparing a system against other methods of recording skills and competencies. It details the design of a prototype digital badging system with military-specific requirements to compare against current Department of Defense (DoD) talent management practices and an industry-standard credentialing system. This research proposes potential use cases for digital badging in a military context, especially in their connection to aligning DoD Cyber Workforce Framework (DCWF) roles to cyber competencies, as an example for how the technology can be used (Department of Defense, 2023). This approach supports efforts to enhance force development and formal and informal learning, especially as large language models (LLMs) increasingly influence learning environments (AFIT Faculty Learning Community, 2025). Badges could potentially provide educators with a trustworthy method of attesting competency. The ability of service members and veterans to represent their skills within the military and in civilian careers is shown through this research's proposed use cases for digital badging. While this study was driven by military need, the approach and process are generalizable and therefore may be applicable to non-military organizations.

Purpose

This research supports DoD priorities by investigating potential improvements for talent management, increasing operational effectiveness. By working with the sponsor of this study, Air Education and Training Command's Learning Professionals Branch (AETC/A3JB), this research is designed to benefit the joint military community by improving how the military develops and manages its members. As shown through efforts to bolster implementation of the DoD Cyber Workforce Framework and the recent addition of Air Force Cyber Warrant Officers, a prime example of the need to manage highly technical positions is the Air and Space Force Cyber community (Defense, 2023; Air Force Public Affairs, 2024). This project's methodology leads to recommendations that answer the following research question (RQ):

RQ1: What is the value proposition of digital badges in the military?

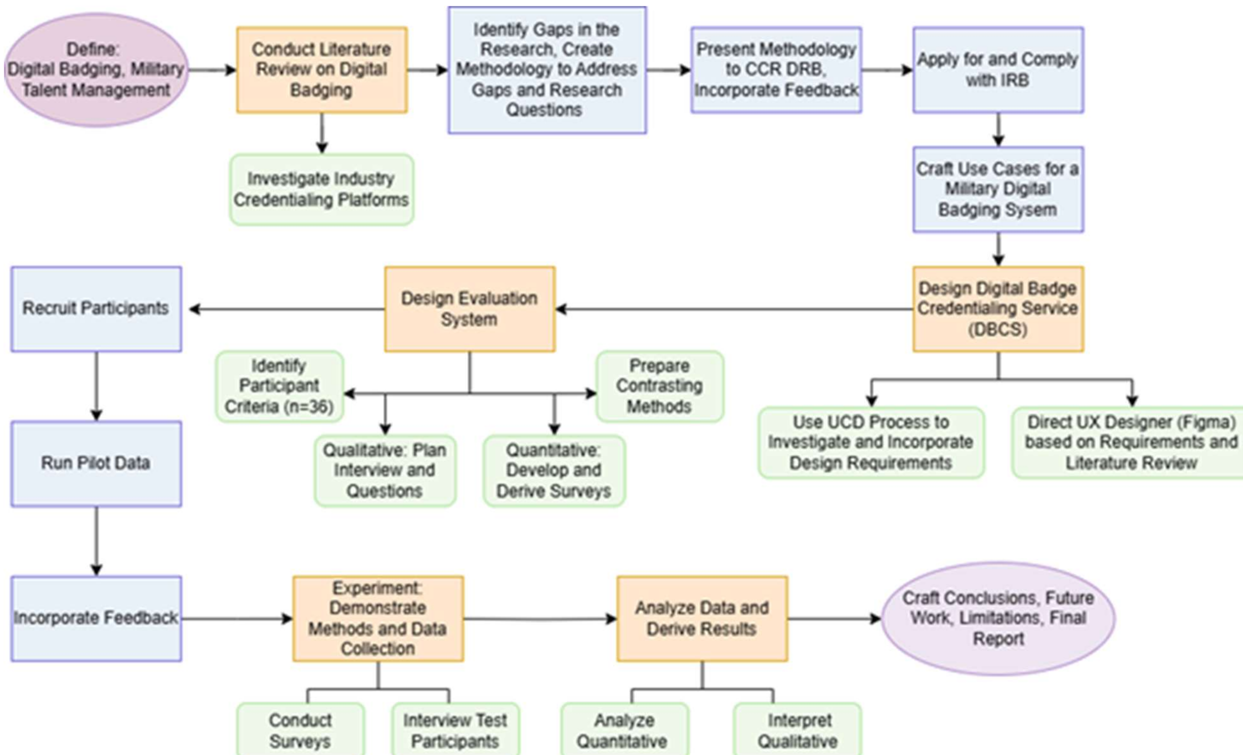

Figure 1. This figure illustrates the research methodology. The methodology starts from the upper left and proceeds sequentially following the arrows. Orange rectangles represent major milestones with supportive green rectangle sub-tasks. Blue rectangles represent intermediate objectives. Purple circles represent the start and end of the research.

The graphic in Figure 1 shows the overall methodology of this research. Beginning at the upper left corner of the diagram, the ``Define" item represents the clarification of the purpose, scope, and objectives of this research. The researchers define the problem domain of this research as military talent management, with digital badges (or micro-credentials) as the proposed solution. From this scope, this study reviews the literature on digital badges, their uses, and perspectives.

Table 1. Addressing Gaps in the Research

Gap	**Addresses Gap**
Gap 1: An incomplete value proposition for military use of digital badges.	Prototype that can be tested for perceived value and allows accurate and realistic evaluation of digital badging against other talent management methods.
Gap 2: Lack of design guidance for digital badging systems for military use.	Prototype can be tested for usability and can receive feedback on usability and design principles.
Gap 3: A lack of demonstrable military micro-credentialing systems.	Prototype can be shown as a proof of concept for digital badging, supported by codified use cases.
Gap 4: Lack of codified use cases for a digital badging system in the military.	Codified use cases for a digital badging system are presented to and evaluated by service members.

Through a synthesis of the literature, researchers find how badging has been successfully implemented in various ways in industry and education (Čubrić and Čubrić, 2016; L. Facey-Shaw et al., 2018; J. Fanfarelli and McDaniel, 2015; Falkner and Falkner, 2014; Fischer and Stabauer, 2022; Lim et al,. 2018; McDaniel, 2016; Menasalvas et al., 2019; Pitt et al., 2019; Rughinis, 2013; Shahriar et al., 2016). However, the value proposition of badging in a military context has yet to be fully explored. Researchers synthesize the gaps in Table 1, and address them through this research, beginning with a description of the badging prototype design in the next section.

Methodology

To compare digital badging to other methods of recording skills and competencies, researchers design a digital badging system prototype. Through a review of the literature and market research on commercial credentialing systems, this team revealed several gaps that make the creation of a military-centered digital badging system desired (Accredible, personal communication, May 20, 2024; Credly, personal communication, May 20, 2024; Instructure, personal communication, May 16, 2024; Sertifier, personal communication, May 14, 2024; Verified Ed, personal communication, May 16, 2024; Virtual Badge, personal communication, May 16, 2024). Commercial digital badging systems primarily target civilian organizations interested in growth, engagement, and earner retention and motivation (J. Fanfarelli and McDaniel, 2015; Gregg et al., 2022; Urfan et al., 2022; Wallis and Martinez, 2013). While these benefits of digital badges are notable, the potential use cases for a military digital badging system may differ from those of a commercial company. The military may be less interested in the user engagement digital badges bring, as military organizations may choose to focus on the granular data that enables talent management. Retention and engagement is not an issue, as military members are not being recruited through badging platforms. Instead, identifying member competencies through aggregate badging data could enable more effective talent placement decisions.

Using a User-Centered Design (UCD) process, researchers work in partnership with AETC/A3JB and a User Interface/User Experience (UI/UX) design team to create a prototype digital badging system, named the Digital Badge Credentialing Service (DBCS) that can convey potential military-specific digital badging use cases, filling a gap in the existing research. These use cases show the value of digital badges for military application and enable evaluation of digital badging as a method for recording skills and competencies.

To answer the research questions, this experiment uses a mixed-methods approach. Researchers gather quantitative data from questionnaires to assess scale, then examine the deeper reasoning behind survey responses in an interview yielding qualitative data. Researchers recruit participants for this experiment with the following criteria:

Researchers recruit participants for this experiment with the following criteria:

a) The subject population includes military members, government civilians, and contractors.

b) The researchers make a special effort to recruit subjects within or related to the cyber career field.

c) The researchers may recruit people with cyber experience from non-DoD organizations based on availability.

The recruitment locations are military organizations. The participants are recruited through email and in-person contact. Participants agree to be part of the experiment, and a meeting is arranged. The experiment occurs over Microsoft Teams, using Teams' recording transcript feature for the interview. After the experiment begins, the protocol in Figure 2 is followed:

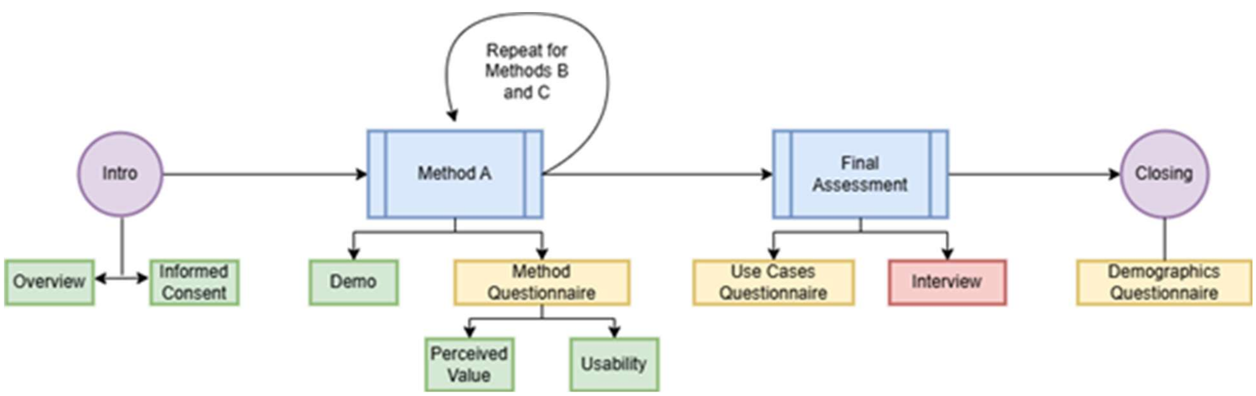

Figure 2. Experiment protocol, showing the flow of an hour-long interview with participants to answer the research questions. Blue rectangles represent major sections of the experiment. Yellow rectangles represent surveys, green rectangles represent supportive sub-items, and red

rectangles are other assessments besides surveys. Purple circles mark the start and end of the experiment.

The estimated time for this protocol is one hour per participant. Researchers replicate this experiment for 36 replicates to perform sufficient statistical analysis. Beginning at the start of Figure 2, the participant is explained the overview and purpose of the study. The researchers then review an Informed Consent Document with the participant in accordance with the Institutional Review Board (IRB) exemption approved for this study. The experiment is voluntary, and participants give their consent to record the interview portion of the video call.

After the introduction, the participants are shown three different methods of recording skills and competencies. Participants will interact with the methods referenced in Table 2 and complete a questionnaire after interacting with each method. A potential nuisance factor for this experiment is survey fatigue, which occurs when participants begin to lose interest in answering repeated survey questions. Survey fatigue is a known, yet uncontrollable, nuisance factor, so a Completely Randomized Design (CRD) is employed (Montgomery, 2017). The use of a CRD controls for potential survey fatigue. The researchers run a Python script prior to each interview, generating the randomly generated control number for each participant and the order each individual participant views the methods. The methods are described in Table 2.

Table 2. This table displays the three methods participants are shown for the A/B/C test of this experiment, including how each method is presented to participants. These methods are presented to participants in a randomly generated order.

Method	**Description**	**Presentation to Participants**
Method A: Current Practice (SEI)	A traditional method of talent management. Special Experience Identifiers (SEIs) record a certain experience or training an individual has received experience (Department of the Air Force, 2024). This method is a legacy system and relies on pre-defined codes.	Screenshared with participants, showing SEI codes and describing relevant use cases.
Method B: Industry-Standard Digital Badging System	Commercially developed badging platform. In coordination with the company and the IRB, the specific product name is not released to participants or in this research.	Screenshared with participants, showing screenshots of the user walkthroughs of the product.
Method C: Digital Badge Credentialing Service (DBCS)	A custom digital badging system built for military use with AETC requirements to record KSATs.	Participants are shown a video of the full DBCS prototype, then are sent a link to interact

		directly with the Figma wireframe prototype feature.

Special Experience Identifiers (SEIs), Method A for this research, are a current method the Air Force uses for talent management. SEIs are three-character alpha-numeric codes and used to document a specific training or experience. For example, a cyber officer who has a role as a Basic Joint Targeting Analyst and has completed the training as specified by the USCYBERCOM/NSA Job Qualification Requirement, would get the SEI 'O6S,' for 'Basic Joint Targeting Analyst' (Air Force, 2024). Personnelists can use SEIs to track experience and fill billets that require that specific experience (Department of the Air Force, 2024). Participants are shown the source document for SEIs, the Air Force Officer Classification Document (AFOCD), an Airman's Department of the Air Force Learning Record (DAFRL) profile with an SEI, and a comparison of SEIs to digital badging systems. A commercial company developing a digital badging system for government use is also shown to participants. This Industry-Standard Digital Badging System, Method B in this research, allows for direct comparison against the DBCS and allows participants to view another example of the capabilities of digital badges.

After the participants interact with the first method for their randomly generated order, they are sent a link to a questionnaire on Microsoft Forms. The questionnaires given after each method include a custom 5-point Likert questionnaire that tests for perceived value, and questions adapted from the System Usability Score (SUS) and Usability Metric for User Experience (UMUX) to test for usability (Lewis, 2018; Lewis et al., 2013). The SUS was considered for this experiment because of its reputation in the UX industry and validation by other researchers, but the 10 questions per questionnaire, multiplied by three methods, would have resulted in a burden of 30 questions per participant on usability (Drew et al., 2018; Lewi,s 2018). Therefore, a modified questionnaire is used, using only three questions adopted from the SUS and UMEX. The three usability questions are chosen to represent a wide selection of UX aspects, including user satisfaction, ease of use, and learnability (Lewis, 2018). The perceived value questionnaire is adopted from the use cases developed for the DBCS and has 4 questions. The breakdown of the number of questionnaire questions is shown in Table 3.

Table 3: Breakdown of the number of questionnaire questions in the experiment.

Questionnaire	**Details**
Method Questionnaire (7 total)	4 perceived value questions, 3 usability questions
Use Cases Questionnaire (9 total)	9 use case importance questions
Demographic Questionnaire (4 total	4 demographic questions

After completing the first method (A, B, or C) demonstration and questionnaire, the participants repeat the same steps for the next two methods. After all methods are presented and all Method Questionnaires are complete, the researchers send the participant a Use Cases Questionnaire. The questionnaire asks participants to rate, on a 5-point Likert scale, how important they rate the functionality described in the codified military badging use cases.

Immediately following the participant's completion of the final questionnaire, a semi-structured interview is conducted to explore the reasoning behind participants' questionnaire responses. Using a flexible interview protocol, the researchers focus on responses to Use Case Questionnaire questions (questions 1-3) and Method Questionnaire questions (questions 3, 6). Additional questions to discuss are chosen based on criteria such as extreme ratings, inconsistencies, or relevance. Participants are asked to elaborate on their answers to provide context and examples to clarify their reasoning. Interviews are recorded, transcribed, and coded to identify themes. The interview protocol is listed in Table 4.

Table 4. The interview protocol researchers follow for each participant, ensuring each interview is generally consistent across participants. This is used to gather deeper insights from participants based on their responses in the questionnaires.

Section	**Details**
Introduction	Explain the purpose of the interview: "To gain a deeper understanding of participant's answers." Researchers ask for consent to record the interview.
Focused Responses	Use Cases Questionnaire Questions 1-3: Discuss the importance of use cases. Method Questionnaire Question 3: Discuss method's value to the service. Method Questionnaire Question 6: Discuss the usability of each method.
Targeted Question Criteria	Follow up on abnormal responses or topics of particular relevance to the research topic.

After the interview, researchers ask for feedback to the experiment protocol, re-inform them of their rights under the IRB, and thank participants for their time. Participants take a final Demographics Questionnaire as the interview concludes. Collected data is analyzed using the methods described in the next section.

Results and Analysis

Using the methodology described, this research conducts 37 interview sessions. Of note, one interview is completed with 4 participants at once in order to meet their scheduling constraints. Additionally, one interview is conducted after qualitative analysis is completed, so only the interview portion of that session is included in this paper, bringing the total to 37 qualitative responses and 36 quantitative responses. The 37th interview is included due to the unique perspective of the participant's Air Force cyber assignment management experience.

The sessions provide unique insight into the perspectives of the service members, government civilians, and contractors interviewed regarding digital badging’s incorporation into military processes. The following sections discuss the quantitative data captured, followed by more perspectives on those findings in the form of qualitative data. Statistical tests and qualitative analysis are performed on the data collected to answer these research questions. There are three major datasets: the Method Questionnaires, the Final Questionnaires, and the interview transcripts. The Method Questionnaires are analyzed first, followed by the Use Cases Questionnaires and interview transcripts. The demographic data collected in the Demographic Questionnaires is used throughout this paper. Finally, the results are discussed further, tying back into the RQ.

First, the systems are compared for value and usability. To discover if there is any difference in perceived value and usability between any of the methods, the Kruskal-Wallis test is used. One-way ANOVA would normally be the correct test to compare a response and one factor with three different levels, however the categorical data collected results in non-normal distributions and a potential lack of homogeneity, violating two out of three of the assumptions of ANOVA (Montgomery 2017). Therefore, the Kruskal-Wallis test is used to answer the RQ. Kruskal-Wallis is well suited to the 5-point Likert scale data collected. Pairwise tests are also conducted to find if there are differences between specific methods. The α= 0.05 for all tests and the hypotheses for these tests are as shown in Table 5. These hypotheses, α value, and rejection criteria are used throughout this paper for all tests. First, Kruskal-Wallis is performed on the perceived value questions. The results of the Kruskal-Wallis test are shown in Table 6.

Table 5. Null and Alternative Hypotheses.

Null Hypothesis(H_0)	The responses are comparable across levels of the factor.
Alternative Hypothesis (H_1)	The responses are not comparable across levels of the factor.

$$\alpha = 0.05$$

Reject the null hypothesis (H0) if $p < \alpha$

Fail to reject the null hypothesis if $p \geq \alpha$

Table 6. Chi Square Approximation using Kruskal-Wallis on Perceived Value and Usability Scores

Measure	χ^2	Probability > χ^2 (p-Value)
Perceived Value	90.4593	<0.0001

Usability	98.0463	< 0.0001

As seen in Table 6, the p-Values listed are all less than the α of 0.05. Therefore, the null hypothesis that the responses are comparable across all levels of the factor is rejected. To determine where the differences between the methods, pairwise testing is done using the nonparametric Steel-Dwass test (InfluentialPoints, n.d.).

Table 7. Nonparametric Comparisons for All Pairs Using Steel-Dwass Method on Perceived Value and Usability Scores.

Level	**-Level**	**Score Mean Difference (p-Value)**
	Perceived Value	
C	A	84.40278 (< .0001)
B	A	61.93056 (< .0001)
C	B	28.44444 (0.0048)
	Usability	
C	A	75.84259 (< .0001)
B	A	58.06481 (< .0001)
C	B	24.35185 (0.0070)

Table 7 shows the test results on perceived value and usability of the three methods, A, B, and C, using the Steel-Dwass test. The three methods are compared as levels against each other in pairs. The same null and alternative hypotheses from Table 5 continue to these tests, the difference being that the responses of only two levels (methods) out of A, B, or C are being compared at once.

For the tests on perceived value, the p-Values for comparing both B and C to A are all $<$ $.0001$. Because they are compared to $\alpha = 0.05$, and $0.05 > 0.0001$, this study rejects the null hypothesis that the responses of the perceived value scores are comparable for B to A and C to A. There was a statistically significant difference between A and B and A and C. Using the Score Mean Differences provided, the Industry-Standard Digital Badging System (B) was perceived as more valuable compared to SEIs (A), and the DBCS (C) was also perceived as more valuable as compared to SEIs (A). That the two badging systems were perceived as more valuable can be suggested due to standard practices in nonparametric statistics. A positive Score Mean Difference suggests that the "Level" group received higher scores on perceived value than the "-Level" group (Washington, n.d.). Therefore, the Industry-Standard Digital Badging System (B) and the DBCS (C) are both perceived to be more valuable than SEIs (A). Additionally, the p-Value for the test between B and C is 0.0048, which is less than the α. Using the positive Score Mean Difference, this suggests that the DBCS is perceived as more valuable than the Industry-Standard.

Similar results are discovered for usability. Using the p-Values in Table 7 and $\alpha = 0.05$, the null hypothesis is rejected for all tests. Method C is suggested to be more usable than Method A or B, and Method B is suggested to be more usable than Method A. Overall, digital badging systems fare well in these tests, and participants find them more valuable and usable than the current method.

It is important to note the reasoning behind why participants rate the two badging systems the way they do in this study. While the questionnaire responses show a slight preference for the DBCS, individual feedback reflects many positive aspects of the Industry-Standard, including a single-screen view of badge criteria and evidence for review, an “intuitive and user-friendly design," a well-organized layout, and the wallet feature. However, one feature that stands out in the DBCS were consumer-focused dashboards. The closest idea participants see in the Industry-Standard system is an “Organization Admin dashboard." While seeing badge issuance data is important, a more critical feature for consumers and decision-makers may be the ability to view badge holder data and make informed decisions based on that information. Participants appreciate dashboards in the DBCS that show conceptual mock-ups of skill gap analysis and “mission readiness vs actual readiness.” Additionally, while neither system has this functionality displayed, a top priority feature participants mentioned was the ability to search, filter, and quickly find member badges and skill-sets. For example, “we need to find a person with X certification or skill - how do we do that quickly with this system?"

Other elements that participants appreciate about the DBCS include earning criteria related to the DoD cyber workforce (appealing to the sample population). They also appreciate the badge creation design. While the Industry-Standard features a clean layout, the more “certificate" layout and design was less appealing to certain users. The DBCS looks more akin to a classic badging system, with a crest-like badge design creator and badges. Participants also like the idea of “recommended badges" for participants in the DBCS based on badges they’ve already earned or their interests. Both systems have aspects of their UI/UX that participants appreciated. Some features that participants discuss are color schemes and simplicity. Interestingly, many participants mention they like the color schemes, layouts, or simplicity of the Industry-Standard, and others have the same comments for the DBCS. Some participants like the more block-like and colorful feel of the DBCS, and some like the more straightforward feel of the Industry-Standard. This suggests that these preferences may be subjective and would require large scale testing to suggest that one of these elements is universally better than another.

The Method Questionnaire questions are created specifically for this experiment. To ensure that the groups of Method Questionnaire questions (1-4) and (5-7) test for the same response variable, their internal consistency is evaluated using Cronbach’s Alpha.

$$\text{Equation 1: } \alpha = \frac{N\bar{c}}{\bar{v}+(N-1)\bar{c}}$$

Equation 1 is used to calculate the coefficient of consistency, using N as the number of items, $\bar{c}$ as the average inter-item covariance among the questions, and $\bar{v}$ as the average variance. As the average inter-item correlation increases, signifying a consistent response, Cronbach’s Alpha increases (Digital Research and Education, n.d.). The output of this equation is the reliability coefficient which suggests the level of internal consistency among the two groups of questions; a reliability coefficient of greater than or equal to 0.70 is broadly considered the acceptable standard (Digital Research and Education, n.d.). Using questionnaire data formatted for multivariate testing, the four perceived value questions and three usability questions are combined and assessed for consistency in Table 8. The Cronbach's Alpha for both sets of questions return higher than 0.70, validating the internal consistency of the question sets.

Table 8. Cronbach's Alpha for Usability and Perceived Value

Measure	Cronbach's Alpha
Usability	0.8790
Perceived Value	0.8627

To better assess the value proposition of digital badging for use in military talent management, codified use cases are created for this paper. These enable study participants, and the Service at large, to better understand how this technology can be used. These use cases are a novel introduction to military badging research. The specific use cases are shown below, enumerated 1-9. To gauge participant feedback on these use cases, they are placed in the Final Questionnaire. Participants are asked to rate how important they believe the functionality each use case would be. To whom it is important to – the Service, themselves, etc. – is intentionally left blank, and elaborated upon in the interview. The results of the Final Questionnaire are 5-point Likert scale data ranging from Strongly Disagree (1) to Strongly Agree (5). To analyze the results of the use cases data, descriptive statistics are calculated to show how important the participants found each use case to be. The use cases are ranked based on their means to give an estimate for how important each functionality was perceived by the participants. While means can provide issues for statistical tests with nominal Likert scale data, the rank order of the means provides a way to show the order in which participants thought the use cases are important (Montgomery 2017).

Table 9. Rank order of use case functionality importance based on participant responses. The Rank column shows the rank order of the use cases, summarized in the middle column. The sorted means are shown to the right, along with the corresponding variances.

Rank	Use Case Summary	Mean	Variance
1	Decision-maker uses badges to find qualified personnel	4.53	0.37
2	Collaboration across units using badges to create cross-functional teams	4.44	0.43
3	Cross-functional skills and informal learning recognized for eligibility	4.33	0.74

4	Identify skill gaps and adjust training or personnel accordingly	4.31	0.39
5	Incentivize technical learning with badges for career growth	4.19	0.96
6	Service member records skills for transition to second career	4.17	0.49
7	Service member records skills to represent themselves to leadership	4.14	1.04
8	Track progress toward badges with feedback for development	4.06	0.63
9	Personalized training recommendations and mentorship alignment	4.00	0.97

Table 9 shows that using badges to find qualified personnel for a technical assignment receives the highest responses from participants. This correlates with participant responses in the interviews. Using badges to create teams and recognize informal learning are also ranked highly by participants. Participants rank using badges to track feedback or as personalized recommendations for mentorship as the least important use cases.

To discover potential correlations between participants' responses and their prior experiences, the Wilcoxon/Kruskal-Wallis test is conducted on all of the tests described in this study. The Demographic Questionnaire collects data on rank, career field, prior experience with digital badges, and prior talent management experience. Using an α of 0.05, no correlation is found between responses and any demographic information.

Finally, interviews with participants are transcribed, summarized, and coded for key points. These annotated transcripts are emailed to participants after the experiment. Participants validate the accuracy and interpretation of the transcripts. Researchers then compile the codified key points and sort into thematic groups (Pan et al. 2024). These thematic groups are shown in the codebook in Table 10.

Theme	**Definition**
Finding Qualified People	Digital badges could improve assignment matching and expedite talent identification, as long as there are clear position requirements.
Skill Gaps	Participants see potential in badges for skill gap analysis, especially in large teams or during leadership transitions, but emphasize the need for widespread adoption.

Member Representation	Badges provide a more objective and verifiable way to represent skills. Their effectiveness for transitioning to a civilian career depends on universal recognition.
SEIs vs Digital Badging Systems	SEIs are perceived as outdated and under used, yet still relevant in some contexts. Digital badges offer needed granularity in tracking achievement.
Design Elements	Participants appreciate consumer-focused dashboards and recommended badge pathways. They encourage system simplicity and additional features like mobile applications and onboarding tutorials.
Implementation Guidance	Concerns include keeping track of another system and system findability. Participants recommend integration with existing platforms, tiered badge structures, review workflows, calculating system burden, and effective messaging.
Badge Assessments	Participants mention concerns over “cheap" badges diluting value, leading to “badge collectors," and failing to reflect actual skills. Badges need effective assessments. Finally, a lack of a badge does not necessarily mean lack of a skill.
Applicability to Cyber Community	Digital badging aligns well with the diverse skillsets and evolving requirements of cyber operations.

Table 10 provides insight into participant perspectives of digital badging. The themes discussed in the interview reflect the positions of the thirty-seven participants, who generally respond positively to questions about the value proposition of badges in the military and their potential use cases. Participants appreciate the granularity, speed, objectivity, and trust badges offer. Many participants believe digital badging could improve an organization’s talent management position matching. However, some participants warn of limitations and advise caution when exploring badging’s potential use cases. Badging systems require effective means of assessment and must consider the potential burdens of cost and time to service members.

Conclusion

This research investigates the value proposition of a digital badging system prototype as a means for the military to record and represent service member achievement. This experimental design fills gaps in the research. A novel methodology for evaluating a digital badging system is proposed and an experiment is conducted using that methodology. Finally, the results of the experiment are shown and discussed, in relation to the research question:

RQ: What is the value proposition of digital badges in the military?

The results of the experiment suggest that digital badging could benefit military talent management processes and provide helpful, granular data for decision-makers. Digital badging systems are seen as more valuable and usable compared to traditional methods of talent management. Use cases specify how digital badging might be used, and participant feedback suggests that the use cases are valid propositions for how badging can work in the military. Overall, participant feedback suggests a growing demand for digital badging technology in the Service.

Despite this feedback, this study has limitations in the sample of participants and the medium through which participants interacted with the methods. More experienced participants in talent management would improve the reputability of this work. Additionally, allowing participants to use the methods in longer and in-depth ways would help solidify their opinions of the methods. Due to the scope of this RQ, future research is needed to fully explore the question of value. A fully functioning military badging system including developing algorithms, search and filtering features, Personalized Learning Paths, and implementation with other enterprise services is needed to understand the practical value of badging. A holistic understanding of the costs, benefits, and challenges will further progress the state of the art. Finally, using digital badges as trustworthy records of accomplishment in LLM learning environments should be further explored.

Declarations

The views expressed in this paper are those of the authors and do not reflect the official policy or position of the United States Air Force, the Department of Defense or the United States Government.

Statement on Artificial Intelligence

By submitting this manuscript, the authors confirm that we have adhered to appropriate ethical guidelines or approvals for the human subjects research conducted in this study, including taking into consideration any federal requirements and institutional guidelines concerning the use of AI in human subjects research.

References

AFIT Faculty Learning Community. 2025. *AFIT Generative AI Teaching Guidebook*. Wright-Patterson AFB: AFIT. https://scholar.afit.edu/docs/140.

Air Force, Department of the. 2024. “Air Force Officer Classification Directory (AFOCD).”

Air Force Public Affairs, Secretary of the. 2024. "Air Force to Begin Accepting Warrant Officer Applications." https://www.afpc.af.mil/News/Article/3756431/air-force-to-begin-accepting-warrant-officer-applications/.

Chan, J. 2023. "Air Force Digital Badges." Thesis, Air Force Institute of Technology, Wright Patterson Air Force Base, OH.

Čubrić, I. S., and G. Čubrić. 2016. "Creativity, Communication, and Collaboration: Grading with Open Badges." In *2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 764–69. https://doi.org/10.1109/MIPRO.2016.7522243.

Defense, Department of. 2023. "Cyberspace Workforce Qualification and Management Program (DoD Manual 8140.03)." https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodm/814003p.pdf.

Digital Research, Institute for, and Education. n.d. "What Does Cronbach's Alpha Mean?" https://stats.oarc.ucla.edu/spss/faq/what-does-cronbachs-alpha-mean/.

Education, Air, and Learning Professionals Branch Training Command. 2023. "EDCS Business Capability Document."

Falkner, N. J. G., and K. E. Falkner. 2014. "Whither, Badges? Or Wither, Badges!: A Metastudy of Badges in Computer Science Education to Clarify Effects, Significance, and Influence." *Proceedings of the 14th Koli Calling International Conference on Computing Education Research*, 127–35. https://doi.org/10.1145/2674683.2674698.

Fischer, T., and M. Stabauer. 2022. "Digital Badge Revision: The Challenge of Handling Feedback." In *2022 20th International Conference on Information Technology Based Higher Education and Training (ITHET)*, 1–4. https://doi.org/10.1109/ITHET56107.2022.10032042.

Gregg, C. Fenton, J. Park, and M. Handley. 2022. "Exploring the 'Why' of Micro-Credentials and Digital Badges: Engineering Students' Motivations for and Perceived Utility of Learning Outside of Class." In *2022 IEEE Frontiers in Education Conference (FIE)*, 1–7. https://doi.org/10.1109/FIE56618.2022.9962376.

InfluentialPoints. n.d. "Non-Parametric Multiple Comparison Tests." https://influentialpoints.com/Training/nonparametric_multiple_comparisons.htm.

J. Fanfarelli, S. Vie, and R. McDaniel. 2015. "Understanding Digital Badges Through Feedback, Reward, and Narrative: A Multidisciplinary Approach to Building Better Badges in Social Environments." *Communication Design Quarterly* 3 (3): 56–60. https://doi.org/10.1145/2792989.2792998.

L. Facey-Shaw, P. van Rosmalen, M. Specht, and J. Bartley-Bryan. 2018. "Educational Functions and Design of Badge Systems: A Conceptual Literature Review." *IEEE Transactions on Learning Technologies* 11 (4): 536–44. https://doi.org/10.1109/TLT.2017.2773508.

Lewis, James R. 2018. "The System Usability Scale: Past, Present, and Future." *International Journal of Human–Computer Interaction* 34 (7): 577–90. https://doi.org/10.1080/10447318.2018.1455307.

Lewis, James R., Brian S. Utesch, and Deborah E. Maher. 2013. "UMUX-LITE: When There's No Time for the SUS." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2099–2102. CHI '13. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2468356.2468729.

Lim, C. L., P. K. Nair, M. J. Keppell, N. Hassan, and E. Ayub. 2018. "Developing a Framework for the University-Wide Implementation of Micro-Credentials and Digital Badges: A Case Study from a Malaysian Private University." In *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, 1715–19. https://doi.org/10.1109/CompComm.2018.8780706.

Mandy R. Drew, Brooke Falcone, and Wendy L. Baccus. 2018. "What Does the System Usability Scale (SUS) Measure?" In *Design, User Experience, and Usability: Theory and Practice*, edited by A. Marcus and W. Wang, 356–66. Cham: Springer International Publishing.

McDaniel, R. 2016. "A Taxonomy for Digital Badge Design in Medical Technologies." In *2016 IEEE International Conference on Serious Games and Applications for Health (SeGAH)*, 1–8. https://doi.org/10.1109/SeGAH.2016.7586254.

Menasalvas, E., A. M. Moreno, and N. Swoboda. 2019. "A Proposal for Recognizing Skills in Data Science Using Open Badges." In *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education*, 316. https://doi.org/10.1145/3304221.3325561.

Montgomery, D. C. 2017. *Design and Analysis of Experiments*. 9th ed. Hoboken, NJ: Wiley.

Pan, Zilong, Shan Li, Juan Zheng, and Lauren Tillstrom Biegley. 2024. "Impacts of Different Gamified Problem-Solving Integration Approaches on Elementary Math: An Engagement and Metacognitive Knowledge Perspective." *Journal of Research on Technology in Education* 0 (0): 1–28. https://doi.org/10.1080/15391523.2024.2437740.

Patel, Pareena. 2023. "Exploratory Work into the Value Proposition of Air Force Cyber Digital Badges." Wright Patterson Air Force Base, OH: Air Force Institute of Technology.

Pitt, Caitlin, Alex Bell, Eduardo Onofre, and Kristin Davis. 2019. "A Badge, Not a Barrier: Designing for-and Throughout-Digital Badge Implementation." In *Proceedings of the*

2019 CHI Conference on Human Factors in Computing Systems, 1–14. https://doi.org/10.1145/3290605.3300920.

Rughinis, Radu. 2013. "Talkative Objects in Need of Interpretation. Re-Thinking Digital Badges in Education." In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, 2099–2108. https://doi.org/10.1145/2468356.2468729.

Shahriar, Hasan, Sherri Peletsverger, Hisham Zafar, Brian Bailey, and Linda Johnston. 2016. "Digital Badges to Enhance Skills and Preparation for a Career in Cyber Security." In *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, 2:622–23. https://doi.org/10.1109/COMPSAC.2016.97.

Urfan, Faisal, Andika Suciani, and Dwi Indrianto Effendi. 2022. "Utilization of Digital Badges to Improve Learners' Retention in Online Course." In *Proceedings of the 5th International Conference on Learning Innovation and Quality Education*, 1–5. https://doi.org/10.1145/3516875.3516967.

Wallis, Patrick, and Maria S. Martinez. 2013. "Motivating Skill-Based Promotion with Badges." In *Proceedings of the 41st Annual ACM SIGUCCS Conference on User Services*, 175–80. https://doi.org/10.1145/2504776.2504805.

Washington, University of. n.d. "Nonparametric Tests." https://courses.washington.edu/psy524a/_book/nonparametric-tests.html.