Cybersecurity Certificate Selection Process: A Data Envelopment Analysis Approach

W. Brian Lambert
Metropolitan State University of Denver

Janos Fustos
Metropolitan State University of Denver

Abel A. Moreno
Metropolitan State University of Denver

Organizations that face the threat of cyberattacks protect against those threats, in part, by hiring cybersecurity professionals who have both relevant experience and information security/cybersecurity certifications. Cybersecurity certificates vary in many aspects, involve costs, and offer associated potential financial returns. Given the large number of cybersecurity certificates available, the decision of which one(s) to pursue may not necessarily be straightforward. In this paper, we illustrate the use of Data Envelopment Analysis (DEA) for cybersecurity certificate selection. Our analysis identifies six of 18 certifications considered that demonstrate a maximal relative efficiency, and we analyze why the other certifications are relatively inefficient.

Keywords: data envelopment analysis, cybersecurity

BACKGROUND

Organizations increasingly face the threat of cyberattacks, both in frequency and severity. The threats range from independent programmers creating viruses to cause mischief, to organized criminals developing ransomware to extort corporations, to governments creating and funding departments to wage cyber warfare against enemies. Recent examples of successful cyberattacks demonstrate that their consequences can be extremely severe. Examples from news headlines include DDoS attacks by Killnet against major US airport websites and an orchestrated breach of Cisco's local network by the UNC2447 cybercrime gang, Lapsus threat actor group, and Yanluowang ransomware operators (Purplesec, 2022).

The cyberworld is constantly evolving, with new hardware and software being introduced to the market, each with their own set of new vulnerabilities. Consequently, cybersecurity threats are dynamic, as attackers are constantly learning, developing, testing, and executing new techniques to exploit known and new weaknesses, and to bypass safeguards to reach their targets.

Consequent with this increasing threat, the demand for, and value of, cybersecurity professionals is increasing. According to the U.S. Bureau of Labor Statistics (2022), the expected job growth for

Information Security Analysts during the period 2012-2031 is 35%, which is much faster than average. While a bachelor's degree in a computer related field is usually required, often employers value additional professional certifications which are more focused on specific capabilities. There are numerous professional certifications available, which creates a decision problem for individuals in the cybersecurity profession in deciding which certification to pursue.

Deciding on which cybersecurity certification(s) to pursue is clearly challenging for a few reasons. First, there are multiple certifications from which to choose, each bringing specific skills to bear against certain aspects of various cybersecurity threats. For example, a CISM certification includes information security governance, program development and management skills useful against enterprise threats (ISACA, 2022), while a CCSP certification ensures cloud related skills to combat attacks against cloud platforms and infrastructure security, and to mitigate cloud application threats ((ISC)2, 2022). In regard to skills, each individual has different strengths, weaknesses, and areas in which they feel successful at solving problems. Similarly, with regards to threats, individuals may find differing levels of interest, challenge, and satisfaction in tackling and combating those threats. The individuals' assessment of their skills and interests is essential for them to make a quality decision, and pursue the certification that best aligns with their personal objectives

Second, with the uncertainty present in such a dynamic field, a highly useful certification today may be eclipsed in value by future threats and consequent certifications tomorrow. For example, 10 years ago the most highly sought-after certifications were vendor certifications (e.g., Cisco and Microsoft), while today with the introduction of the global APT threats, the more offensive and technical certifications are the "hot tickets" (Messina, 2022) for cybersecurity professionals. Evaluating and incorporating uncertainty into any decision is highly dependent on the individual and their personal preferences with regards to risk.

Thirdly, each certification includes costs and benefits, creating trade-offs between the certifications. Trade-offs are, loosely speaking, increasing the achievement of one objective by sacrificing the achievement of some other objective. For example, pursuing a certificate that generates a higher expected salary may require a higher fee to sit for the exam for that certificate. While this example is of a single trade-off that is relatively clear, assessing the trade-offs across a multitude of factors for many options can become challenging. In this paper we illustrate the use of the Data Envelopment Analysis (DEA) optimization technique to assist individuals in evaluating the trade-offs between the different certifications.

DATA ENVELOPMENT ANALYSIS (DEA) EFFICIENCY MODEL

Data Envelopment Analysis (DEA) is an optimization technique developed by Charnes, et al. (1978) that produces a single measure of relative efficiency between multiple "decision-making units (DMU)." The literature includes a plethora of articles on the use of DEA, with comprehensive reviews and surveys of modeling techniques ((Cook, et al., 2009) and (Kao, 2014)) and applications specific to industries ((Kaffash, et al., 2020) for insurance and (Paradi, 2014) for banking), for example.

However, the literature appears to be in its infancy regarding applications of DEA in the cyber arena. Nguyen, et al. (2022) use DEA to evaluate the production efficiency of cybersecurity firms. Voronenko, et al. (2022) use DEA to analyze various European countries' efficiencies in the context of cybersecurity, and then use that analysis to identify main vulnerabilities in the specific country of Ukraine.

In this paper, the DMUs are various cybersecurity certifications, all of which are characterized by a common set of quantitative attributes for inputs and outputs. The DEA efficiency model produces a single measure reflective of the relative efficiency by which each DMU transforms these inputs into outputs. This relative efficiency is one measure of the various trade-offs, allowing a rank-ordering of the certifications. This ordering then may assist individuals in selecting which certifications to pursue.

TABLE 1
INPUT AND OUTPUT ATTRIBUTES FOR CYBERSECURITY CERTIFICATIONS

		Inputs			Outputs	
Unit	Certification	Preparation Time (Months)	Experience Required or Recommended (Years)	Exam Fee ($)	Demand (Job Postings)	Salary ($k)
1	CCNA Security	3	1	400	10,009	89
2	Network+	3	1	358	7,817	71
3	Security+	3	2	392	11,981	79
4	CCNP Security	6	3	400	1,814	113
5	CEH	6	2	1,050	3,072	83
6	CISSP	9	5	749	11,981	121
7	SSCP	9	1	249	2,323	78
8	CISM	6	5	760	4,818	131
9	CISA	4	5	760	8,718	108
10	CCSP	4	5	599	2,635	76
11	GSEC	4	4	949	3,393	102
12	GPEN	3	3	949	750	104
13	ECSA	6	2	514	222	83
14	CRISC	6	3	760	1,593	132
15	GCIH	3	2	2,499	2,458	100
16	OSCP	6	2	850	1,583	96
17	CASP	6	5	494	1,289	95
18	CySA	3	4	392	2,059	70

In addition to the rank-ordered list of DMUs, the DEA model provides information useful for assessing why certain certifications are rated as inefficient compared to those rated as efficient. Specifically, the sensitivity analysis output of the DEA optimization model indicates for each inefficient DMU a reference set of efficient DMUs. The inefficient DMU inputs and outputs are compared to those of the subset of efficient DMUs to understand why they are ranked differently. We present an example of this analysis in the results section of this paper.

Developing a common set of input and output attributes is not trivial. As these are the basis for the relative efficiency measure, each DMU must be fairly characterized by each attribute. For example, including an output (or input) such as geographical demand which is provided by only a subset of DMUs would unfairly penalize the efficiency of DMUs not providing that output. Also, there must be data available for all input and output attributes, for all DMUs. This can be challenging if attributes are chosen which are proprietary to issuing organizations, such as their costs to administer and award certifications.

For this analysis, we selected as inputs three requirements for each certification, and as outputs two benefits of each certification. The inputs included 1) the candidates' required or recommended years of experience for the certification, 2) the estimated exam preparation time in months, and 3) the fee required to sit for the exam in U.S. dollars. The outputs included 1) the average salary of jobs requiring the certification in U.S. dollars from payscale.com, and 2) the demand for individuals with each certification in number of job postings on indeed.com. Table 1 above shows these input and output values for the 18 certifications we evaluated with DEA.

DEA MODEL FORMULATION

Given a set of DMUs, the DEA model is a linear program which calculates an efficiency for a single DMU, relative to the other DMUs. The linear program maximizes the weighted sum of that specific DMU's outputs by varying the weights applied to the inputs and outputs of all the DMUs in the set. Solving this same linear program for each DMU in the set enables a rank-ordering of the DMUs by efficiency. The mathematical formulation of the DEA model follows.

Indices

- j = decision making unit $\in \{1..n\}$
- i = input $\in \{1..m\}$
- r = output $\in \{1..s\}$

Parameters

- y_{rj} = value of output r on unit j
- x_{ij} = value of input i on unit j

Decision Variables

- u_r = weight given to the r^{th} output
- v_i = weight given to the i^{th} input

Objective Function:

$$Max \quad e_j = \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \tag{1}$$

Constraints:

$$\sum_{i=1}^{m} v_i x_{ij} = 1 \tag{2}$$

$$\sum_{r=1}^{s} u_r y_{rj} \leq \sum_{i=1}^{m} v_i x_{ij} \quad \forall j \tag{3}$$

$$u_r, v_i \geq 0 \quad \forall\, r, j \tag{4}$$

The objective function (1) allows the unit being solved for the chance to select those best weights to maximize its efficiency. The denominator of (1) is constrained by (2) to equal one, thereby preventing non-linearities or unbounded solutions. Constraints (3) prohibit any DMU from having an efficiency greater than 100%. Constraints (4) ensure the decision variables remain non-negative.

RESULTS

In general, the DEA model solution includes a slack variable (shadow price) for each of the difference constraints (3) associated with a single DMU. The solution for any inefficient DMU, say unit A, will have a set of non-zero slack variables, for say units B and C, which we refer to as unit A's reference set. The technical interpretation of the reference set is that it contains the coefficients for a linear combination of those units, again B and C, with which to construct a hypothetical unit capable of efficiently transforming unit A's inputs into outputs.

TABLE 2
RELATIVE EFFICIENCIES AND REFERENCE SETS FOR THE 18 CYBERSECURITY CERTIFICATIONS

Unit	Certification	DEA Efficiency	Reference Set
1	CCNA Security	1	
2	Network+	0.8831	1, 3, 7
3	Security+	1	
4	CCNP Security	1	
5	CEH	0.4663	1
6	CISSP	0.6894	1, 3, 7
7	SSCP	1	
8	CISM	0.7638	1, 4
9	CISA	0.8705	1, 12
10	CCSP	0.6309	1, 12
11	GSEC	0.7845	1, 12
12	GPEN	1	
13	ECSA	0.6593	1, 4, 7
14	CRISC	0.7697	1, 4
15	GCIH	1	
16	OSCP	0.5393	1
17	CASP	0.7406	1, 4
18	CySA	0.7982	1, 4

In our specific application we cannot construct a hypothetical certification. However, examining the units included within the reference set provides insights as to why the inefficient certification was not evaluated as relatively efficient compared to the other certifications. Table 2 includes the DEA relative efficiency for all of the 18 certifications, and the reference set for those certifications evaluated as relatively inefficient.

Consider the case of the relatively inefficient certification CEH, which has a single efficient certification, CCNA Security, in its reference set. Table 3 displays an extract from Table 1, including input and output data only for these two certifications. Each of the inputs for the CEH certification is at least twice the amount as those inputs for the CCNA Security certification, while both of the outputs for CEH are below those for CCNA Security, with demand being less than one third. Here the CCNA Security certification dominates the CEH certification, where based on these factors, the preferred alternative is unambiguous. However, note that most of the reference sets of inefficient certifications include more than one efficient certification.

Consider the case of the relatively inefficient CISSP certification and its reference set of three efficient certifications: CCNA Security, Security+, and SSCP. Table 4 displays an extract from Table 1, including input and output data for these certifications. Given its high outputs, CISSP is not clearly dominated by any single certification in its reference set. Instead, it is the combination of input and output values, i.e., what inputs does it take to achieve those outputs, that makes CISSP relatively inefficient compared to the others. In relation to the CCNA Security certification, the CISSP requires 3x as much exam preparation, 5x the experience, and roughly 2x the exam entrance fee. In relation to the Security+ certification, the CISSP requires 3x the exam preparation, 2.5x the experience, and 2x the exam entrance fee. Finally, in relation to the SSCP certification, the CISSP requires 5x the experience and 3x the exam entrance fee.

TABLE 3
INPUT AND OUTPUT ATTRIBUTES FOR THE INEFFICIENT CEH CERTIFICATION AND THE CCNA SECURITY CERTIFICATION IN ITS REFERENCE SET

		Inputs			Outputs	
Unit	Certification	Preparation Time (Months)	Experience Required or Recommended (Years)	Exam Fee ($)	Demand (Job Postings)	Salary ($k)
5	CEH	6	2	1,050	3,072	83
1	CCNA Security	3	1	400	10,009	89

This example reflects the more common case which lacks domination, but instead includes multiple and mixed trade-offs between alternatives. While the inputs are definitely higher for the CISSP certification than for those in its reference set, the CISSP outputs include the highest demand and third highest salary of all certifications in our data set. If a cybersecurity professional's primary objective is salary and job opportunities, then CISSP would be highly attractive. However, if that same individual has limited experience or time to prepare, other certifications may be more appropriate.

TABLE 4
INPUT AND OUTPUT ATTRIBUTES FOR THE INEFFICIENT CISSP CERTIFICATION AND THOSE CERTIFICATIONS IN ITS REFERENCE SET

		Inputs			Outputs	
Unit	Certification	Preparation Time (Months)	Experience Required or Recommended (Years)	Exam Fee ($)	Demand (Job Postings)	Salary ($k)
6	CISSP	9	5	749	11,981	121
1	CCNA Security	3	1	400	10,009	89
3	Security+	3	2	392	11,981	79
7	SSCP	9	1	249	2,323	78

CONCLUSIONS

Making good quality decisions requires an understanding of the trade-offs, often between multiple factors and many alternatives. We have illustrated in this paper the usefulness of the Data Envelopment Analysis model in supporting the decision about which cybersecurity certification to pursue. The DEA approach identifies which certifications are relatively efficient in their transformation of inputs to outputs. For those certifications identified as inefficient, the DEA solution identifies those specific efficient certifications to examine to better understand the trade-offs which make them inefficient.

REFERENCES

Charnes, A., Cooper, W.W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444.

Cook, W.D., & Seiford, L.M. (2009). Data envelopment analysis (DEA)–thirty years on. *European Journal of Operational Research*, *192*(1), 1–17.

ISACA. (2022). *CISM*. Retrieved December 22, 2022, from https://www.isaca.org/credentialing/cism

(ISC)2. (2022). *CCSP – The Industry's Premier Cloud Security Certification*. Retrieved December 22, 2022, from https://www.isc2.org/Certifications/CCSP

Kaffash, S., Azizi, R., Huang, Y., & Zhu, J. (2020). A survey of data envelopment analysis applications in the insurance industry 1993–2018. *European Journal of Operational Research, 284*(3), 801–813.

Kao, C. (2014). Network data envelopment analysis: A review. *European Journal of Operational Research*, *239*(1), 1–16.

Messina, G. (2023, January 19). *7 top security certifications you should have in 2023.* Infosec Resources. Retrieved from https://resources.infosecinstitute.com/topic/7-top-security-certifications-you-should-have/

Nguyen, V.T., Wang, C.-N, Yang, F.-C., & Vo, T.M.N. (2022). Efficiency evaluation of cyber security based on EBM-DEA model. *The Eurasia Proceedings of Science Technology Engineering and Mathematics, 17*, 38–44.

Paradi, J.C., & Zhu, H. (2013). A survey on bank branch efficiency and performance research with data envelopment analysis. *Omega*, *41*(1), 61–79.

Purplesec. (2022). *Expert Analysis on The Latest Cyber Attacks, August and October Newsletters*. Retrieved on December 22, 2022 from https://purplesec.us/security-insights/data-breaches/

U.S. Bureau of Labor Statistics. (2022, September 8). *Information Security Analysts: Occupational Outlook Handbook:* Retrieved from https://www.bls.gov/ooh/computer-and-information-technology/information-security-analysts.htm

Voronenko, I., Nehrey, M., Laptieva, A., Babenko, V., & Rohoza, K. (2022). National cybersecurity: Assessment, risks and trends. *International Journal of Embedded Systems, 15*(3), 226–238.