

RESEARCH ARTICLE

Tracing institutional change in the officer corps using textual data from a military school: promise, pitfalls, and ethical considerations

Tamir Libel¹  and Krystal Hachey²

¹University of Bamberg, Bamberg, Germany

²Director General Military Personnel Research and Analysis (DGMPPRA), Department of National Defence (DND), Ottawa, Canada

Corresponding author: Tamir Libel; Email: tamir.libel@uni-bamberg.de

Received: 05 September 2024; **Revised:** 05 March 2025; **Accepted:** 28 April 2025

Keywords: institutional change; officership; professional military education; quantitative text analysis; text-as-data

Abstract

In recent decades, researchers have analyzed professional military education (PME) organizations to understand the characteristics and transformation of the core of military culture, the officer corps. Several historical studies have demonstrated the potential of this approach, but they were limited by both theoretical and methodological hurdles. This paper presents a new historical-institutionalist framework for analyzing officership and PME, integrating computational social science methods for large-scale data collection and analysis to overcome limited access to military environments and the intensive manual labor required for data collection and analysis. Furthermore, in an era where direct demographic data are increasingly being removed from the public domain, our indirect estimation methods provide one of the few viable alternatives for tracking institutional change. This approach will be demonstrated using web-scraping and a quantitative text analysis of the entire repository of theses from an elite American military school.

Policy Significance Statement

By employing innovative data collection on a large corpus of student theses, this study reveals two critical insights for policymakers. First, our analysis reveals a persistent lack of demographic diversity within the elite US Army's military school, indicating that diversity efforts have not altered the officer corps despite broader societal trends. Second, we document a significant shift in doctrinal focus over decades, suggesting that while the intellectual content adapts to external pressures, the demographic profile remains stable. This discrepancy underscores the importance of large-scale data for monitoring institutional change. As direct demographic data become increasingly inaccessible due to federal removals, indirect estimation methods, like those employed here, are among the few viable approaches available.

1. Introduction

The officer corps is hegemonic in the armed forces (Crosbie et al., 2019), with clear hierarchical divisions between junior, mid-career, and senior officers. Upward mobility from junior officer ranks up to general

 This research article was awarded Open Materials badge for transparent practices. See the Data Availability Statement for details.

© Crown Copyright - His Majesty the King in Right of Canada and the Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

officers involves periodic attendance at professional military education (PME) organizations, which impart officers with the attitudes, norms, and expert knowledge required for executing their duties (Libel, 2016). PME organizations are sometimes used as proxies to assess the officer corps' sociological profile (Holting and Brænder, 2020) or to analyze military expertise (Libel, 2016). However, this approach is constrained by limited access to direct data and by conceptual challenges in understanding the officer corps as a profession (Davis, 1980; Libel, 2010, 2016; van Creveld, 1990).

This paper outlines a strategy to overcome both the conceptual and data-access issues. First, it builds on Libel's (2021) recent reconceptualization of officership and PME. Second, it traces institutional change in the officer corps using a computational strategy of web-scraping a large-scale research monograph collection from the School of Advanced Military Studies (SAMS), where we infer students' gender¹ and ethnicity using machine learning based on the authors' names, and employ structural topic modeling to trace shifts in the topics discussed in the corpus (i.e., the thesis dataset). In an era when direct demographic data are increasingly removed from the public domain—driven by federal efforts to curtail Diversity, Equity, and Inclusion (DEI) initiatives—indirect estimation methods such as those used here become essential tools for monitoring institutional change.

This study asks: How have the demographic (gender and ethnicity) and doctrinal/ideational dimensions of SAMS changed from 1985 to 2019, and what do these changes reveal about institutional inertia and adaptation within the US Army? Our objectives are threefold: (1) to validate gender inferences using publicly available class size data; (2) to analyze shifts in doctrinal topics via structural topic modeling; and (3) to discuss the broader implications for monitoring institutional change when direct data are increasingly inaccessible.

Finally, this paper provides insights on the way forward, including practical and ethical challenges, and emphasizes the critical importance of indirect estimation methods in tracking institutional change in an era of diminishing direct demographic data availability.

2. Rethinking officership and military education as institutions

Scholars and practitioners of military affairs have long agreed that the culture of the armed forces is important to performance on the battlefield and to relations with the parent society (Libel, 2016; Murray, 2000; Muth, 2011). However, there seems to be no consensus on how to operationalize it. Muth (2011) argued that military culture is constructed by both peer-learning and PME organizations. Following him, Libel (2016) suggested that *analyzing* the social composition and intellectual content of PME organizations can be an important metric for understanding military culture. Further, Libel (2021) has incorporated Murray's (2000, p. 144) definition of military culture as “the cultural patterns by which officers judge themselves and their environment,” equating it with the concepts of officership and the military profession. *In this study, we posit that officership and PME are interrelated institutions², with the core characteristics of officership being continuously constructed and reproduced within PME.*

To trace the evolution of an institution across time, historical institutionalism (Immergut, 2006; Steinmo, 2008) conceives of institutions as formal or informal procedures, routines, norms, and conventions embedded in the organizational structure of political or economic relations (Hall and Taylor, 1996; Steinmo, 2008). The focus tends to be on state-level institutions that structure conflicts and outcomes (Amenta and Ramsey, 2010), and the concepts of *path dependency* and *critical juncture* are important. Regarding the former, the central claim of historical institutionalism is that choices formed when an institution is being established, or when a policy is formulated, have a constraining effect into the future. This dynamic occurs because institutions and policies exhibit inertia; once established, significant effort is required to change their trajectory (Greener, 2005, p. 62).

¹ The binary gender attribution reflects the data sources used—i.e., aggregated information concerning the historical patterns of binary gender attribution of names.

² Following Ebbinghaus (2005, p. 6), an institution is defined as “social rules, norms and ideas (Leitideen) that guide, but also restrain, social behaviour.”

Thus, path dependency supplies the historical dimension of historical institutionalism (Greener, 2005, p. 62), where the longer an institution is locked onto a certain path, the harder it will be for it to break away from that path. Of course, major institutional change does happen; these “critical junctures” are rare moments in which agents have a wider range of policy options than usual, with the selected option having ramifications sufficient to generate new, self-reinforcing path-dependent processes (Capoccia and Kelemen, 2007). For example, the end of the Cold War (circa 1991) and the strategic realignments following the events of 9/11 serve as critical junctures that have driven shifts in military doctrine, even though demographic patterns may remain relatively unchanged. Hence, exploring PME from a historical institutionalist perspective involves analyzing the organizational structures within which the institution is embedded—namely, PME organizations—and how they construct and reproduce officership. Viewing PME institutions through this lens provides three key insights: (1) change (or its absence) can only be assessed over time relative to past conditions; (2) change often originates at a crisis or critical juncture; and (3) such analysis reveals what changes have occurred, though not necessarily why they occurred (Steinmo, 2008).

As an institution, officership (i.e., the military profession), as outlined in Figure 1, can be defined as the expertise, attitudes, ethics, and rules that constitute commissioned command within the armed forces.³ Drawing upon Ebbinghaus (2005), we conceptualize officership as comprising three interrelated components: norms, rules, and ideas. Norms constitute the informal components that frame, enable, and constrain perceptions of who is entitled to be an officer and what constitutes officership, including expectations of behavior. Rules are the formal components that govern the recruitment, selection, and promotion of officers. Ideas include both formal, codified doctrines (e.g., military doctrine) and informal beliefs that shape how officers should behave (an example being the literature on military professionalism; Libel, 2020).

The analysis of officership as an institution largely depends on understanding PME as “an institution which sets the social rules, norms and ideas for the organization tasked with the education of officers” (Libel, 2021, p. 122). Figure 2, drawing again upon Ebbinghaus (2005, p. 6), outlines the components of PME, highlighting its dual nature in which military and academic norms coexist, though they do not always align.

Regarding officership norms, the dual nature of the institution means that PME incorporates both military and academic norms, which do not always accord with each other within the organization (Libel, 2016). PME imparts a variety of informal ideas through faculty-student interactions in the curriculum and through individual and group work on research papers, theses, articles, and books. In parallel, formal ideas are inculcated through military doctrine, which constitutes the core of military expertise. The rules component is defined by the formal policies governing the recruitment and management of staff, faculty, and students, as well as the overall administration of PME.



Figure 1. Officership as an institution⁴.

³ The authors' definition.

⁴ Figure adapted from Libel (2021) based on the definition of institution by Ebbinghaus (2005).

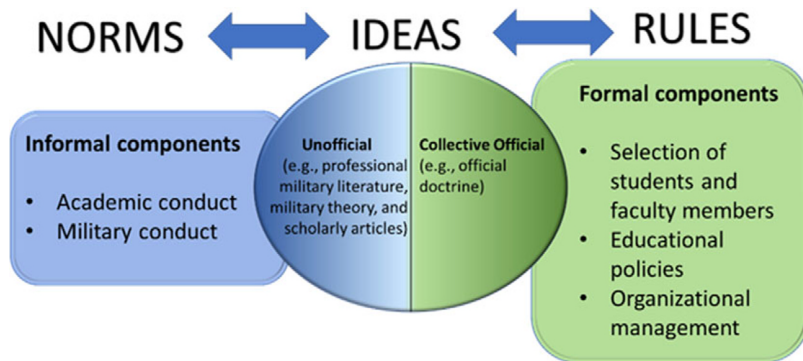


Figure 2. Military education as an institution⁵.

Like other organizations, the military has a defined, official elite—the officer corps. Unlike many other professions, however, officers are obliged to attend specialized professional education courses, which constitute part of the requirements for promotion. Using historical institutionalism as a lens to analyze the different dimensions of PME organizations can shed light on the evolution of PME and, by extension, officership as an institution. For example, analyzing changes in the demographics (e.g., gender or ethnicity) of a senior military education school’s student body may indicate shifts in institutional norms or rules. While the social dimension includes many facets (Segal, 2004), for the purposes of this article, we focus on gender and ethnicity. Additionally, Western PME requires graduate students to produce research papers that often align with the interests of senior headquarters (Libel, 2010); analyzing these documents can help trace the development of the institution’s ideational character over time.

Accordingly, the analytical strategy outlined here focuses on exploring the social and ideational dimensions of PME to assess whether change has occurred. A longitudinal study of any PME requires extensive data on military settings, which are often unavailable due to ethical and legal constraints (Government of Canada, 2019), even when efforts have been made to provide open access (Government of Canada, 2022). Furthermore, analyzing large qualitative datasets can be labor-intensive. Nonetheless, recent positive developments—such as increased publicly available web-based military data and advances in text-as-data methodologies (Grimmer et al., 2022)—offer promising alternatives for studying institutional change, particularly when direct demographic data are no longer accessible.

Theory-based hypotheses: Based on the above theoretical framework, we hypothesize that: (1) due to institutional inertia, the demographic composition (gender and ethnicity) of senior military education remains largely stable over time; and (2) critical junctures—such as the end of the Cold War and post-9/11 strategic shifts—will result in significant changes in the doctrinal or ideational content of PME, even if demographic patterns remain unchanged.

3. Aim

This paper aims to show the utility of quantitative text analysis for capitalizing on the ever-expanding textual data from armed forces to trace changes in the social and ideational dimensions of PME. We focus on one military education organization, the US Army’s School of Advanced Military Studies (SAMS). Founded in the early 1980s, SAMS emerged as an elite institution dedicated to producing innovative operational planners. Its highly selective admissions process ensures that only the top-performing officers are admitted, resulting in relatively small class sizes (Benson, 2010). Moreover, SAMS’ rigorous, research-oriented curriculum is designed to respond to high-level strategic directives from headquarters, and has played a pivotal role in shaping modern US military doctrine (Benson, 2010; Libel, 2010). This

⁵ Figure adapted from Libel (2021) based on the definition of the institution by Ebbinghaus (2005).

institutional background underscores SAMS' significance as a case study for exploring both enduring demographic patterns and evolving doctrinal perspectives.

This research output from SAMS provides the data for the current paper. An automated data collection method, web-scraping, was used to collect all the research output of SAMS students between 1985 and 2019. Although the metadata do not include detailed demographic identifiers, we employ machine learning algorithms to infer gender and ethnicity from the available author information. The textual data are analyzed inductively using a non-supervised machine learning algorithm, structural topic model (STM), to trace the evolution of the social (i.e., ethnicity and gender) and ideational dimensions of SAMS.

4. Methodology

The following sections describe, evaluate, and discuss the limitations of the data collection and analysis. We also detail the text pre-processing steps, note that replication scripts are available on our OSF repository, and address ethical considerations regarding the use of publicly available data.

4.1. Data collection

Data were collected in July 2020 using a web-scraping script written in the R statistical programming language for the US Army Command and General Staff College's Ike Skelton Combined Arms Research Library (CARL) Digital Library.⁶ We obtained all SAMS students' research monographs from 1985 to 2019, yielding an initial corpus of 3948 items (i.e., research monographs and their metadata). Four duplicate items were identified and removed, resulting in a final corpus of 3944 items.

The data preparation workflow followed the recommendations of Welbers et al. (2017): importing text, performing string operations, pre-processing, creating a document-term matrix, filtering, and weighting. Weighting was not implemented because it offered no analytical advantage in this study. The resulting dataset, containing both the metadata and full text of each monograph, served as the data source for the topic-modeling analysis. This dataset also provided the basis for generating a dataset of the research monographs' authors. As some authors contributed more than one monograph, their full names were formatted to exclude duplication. The resulting author dataset contained 3052 observations and was used for the gender and ethnicity inferences. In addition, a version of the dataset with a variable indicating the year in which the respective author published their monograph was produced, yielding a final dataset of 3710 observations.

All data used in this study are publicly available from the CARL Digital Library, and no private or sensitive personal information was collected. The web-scraping process complied with the library's robots.txt guidelines. Furthermore, full replication scripts, detailed pre-processing steps (including tokenization, stopword removal, and stemming), and other technical documentation are available on our OSF repository, ensuring transparency and replicability of our research.

4.2. Methodology

The analysis was divided into two parts. First, an inference of SAMS gender and ethnic composition per academic class was conducted on the last dataset mentioned in the previous section, which includes the variables author's full name, year (i.e., year of graduation), gender1, gender2, and ethnicity. The graduation year of a SAMS class was considered the year of the class (e.g., academic year 2004–2005 is considered the 2005 class).

It should be noted that in 2007 SAMS was directed to conduct two classes per year of its main programme, the Advanced Military Studies Program (AMSP). The first year in which two classes graduated was 2008, with courses running from January through December. Two annual intakes

⁶ The automated data collection was conducted in full alignment to the website's rules. These are specified in the following webpage: <https://cgsc.contentdm.oclc.org/robots.txt>.

continued until 2014, with the second 2014 course graduating in December 2014 (K. Benson, personal communication, March 14, 2023). This accounted for a higher number of SAMS graduates for the 2008–2014 period. As the authors’ names were grouped by year of publication (i.e., graduation), the increase in the number of SAMS classes does not influence the composition of the dataset beyond reflecting a higher number of graduates during that period.

Second, an unsupervised learning algorithm was used to analyze the ideational content of students’ research monographs. Combined, these analyses served as proxies for tracing the evolution of the social and ideational components of US Army PME and, by implication, officership in the US Army.

The gender and ethnic identities of the monographs’ authors were inferred using machine learning algorithms based on the author’s first name for gender and full name for ethnicity. Three algorithms were used via R packages: two for gender (the “gender” package (Blevins and Mullen, 2015) and GenderGuesser (Coccopuffs, 2020)), and one for ethnicity (rethnicity [Xie, 2021]). A major hurdle in inferring gender from the author’s name is the changing patterns in gender association over time (Blevins and Mullen, 2015). As SAMS students were born mostly in the 1950s⁷ onward, such fluctuations were minimal and of negligible impact on the current study.

The gender package contains a predictive algorithm that relies on several datasets concerning first-name gender attributions in the United States., most notably from the American Social Security Administration (SSA). As most SAMS students are US citizens, the SSA dataset was used to infer authors’ gender based on their first names (Blevins and Mullen, 2015). While the gender package relies on official data, the GenderGuesser package uses a commercial API (genderize.io) providing access to over 6 million validated names.

For each algorithm, the inferred gender was saved as a new variable, adding two variables to the author dataset. The author’s ethnicity was inferred from their full name rather than their first name, as the latter lacks sufficient information for ethnicity estimation. This inference was conducted using the R package rethnicity (Xie, 2021), based on the Florida Voter Registration dataset. As most monograph authors are American citizens, the US-centered data was deemed appropriate. The resulting inferences were coded into a new ethnicity variable with four categories: Asian, Black, Hispanic, and White. These three algorithms are supervised learning methods, trained on a predetermined dataset to classify observations.

In contrast, the algorithm for the ideational analysis employed a non-supervised learning approach. This approach uses statistical learning to discover latent dimensions in the data. The algorithm used for analyzing the monograph corpus is the STM R package, built upon the latent Dirichlet allocation (LDA) model in which every document contains a mixture of topics, and topics are mixtures of words (Roberts et al., 2016). Unlike most topic modeling algorithms, STM allows exploration of how covariates (e.g., author’s gender) influence topic distributions among documents.

Text pre-processing for topic modeling included standard procedures such as tokenization, removal of stopwords, and stemming, as detailed in Section 4.1. We experimented with various values for the number of topics ($k = 10, 20, 50$) using the STM R package. Diagnostic metrics—such as semantic coherence, heldout likelihood, and residuals—were computed to assess model quality. Although initial diagnostics suggested that a 20-topic model might be optimal, further visual inspection of the topic distributions indicated that a 50-topic model provided richer interpretability and more nuanced thematic distinctions. Full details of the diagnostic metrics and the model selection process are available in the technical appendix on our OSF repository.

5. Results

Predictive algorithms need to be assessed for reliability and validity. A two-step inter-coder reliability approach was used to assess the gender algorithms. First, we hand-coded the first names of the authors

⁷ This assumption is based on the common knowledge about the average age of SAMS’ attendants given their ranks, and the founding year of SAMS.

using three possible codes: 0 for male, 1 for female, and not applicable (NA) for missing data. The intercoder reliability between their results was evaluated using Cohen's kappa unweighted test, with the result being 1, indicating complete agreement. Second, the results of each of the predictive algorithms were compared with the hand-coded ones using the same test, as well as to each other.

While the Cohen's kappa result for the inter-coder reliability test for the individual hand coders was 1.00, the result of the test for agreement between the human coders and variable gender1 (the inferences from the gender package) was 0.856, indicating strong agreement. The result of the test for agreement between the human coders and variable gender2 (the inferences from GenderGuesser) was 0.0866, indicating no agreement.⁸ Finally, a test was conducted for agreement between variables gender1 and gender2, resulting in 0.0873, indicating no agreement. Agreement between human coders—considered the gold standard for algorithms (McHugh, 2012; Song et al., 2020; He and Schonlau, 2021)—is the maximum that could be obtained, so the results for gender1 should be interpreted as reliable. Considering the low result for gender2, the official data source used for training the gender package provided better inferences than those of GenderGuesser.

To further validate our gender inferences, we compared the predicted class totals (PreTotals) with the observed class totals (ObsTotals) for years for which unambiguous data were available. The following table (Table 1) summarizes this comparison. In the early years, the predicted totals closely match the observed totals, while modest discrepancies are observed in later years. This supports the reliability of our gender inference approach using the gender package.

As displayed in Figures 3 and 4, the overall gender distribution per year is similar across the algorithms. Despite the gradual increase in the total number of students over the years, the proportion of female students to the total number of students per year remained roughly constant. This observation further supports the inference that no significant change occurred in the gender facet of the social dimension of SAMS.

We intended to assess the validity of inferences by comparing the real gender distribution for classes to the inferred gender distribution for each academic year. Unfortunately, only partial SAMS numbers were obtained for comparison, rendering the use of statistical tests for validity impractical. The limited external data available, which are included in the online appendix, suggest that the inferences made by the gender package are more accurate than those made by GenderGuesser.⁹

Figure 5 provides the overall distribution of inferred ethnicity among SAMS students from 1985 to 2019.

Despite the gradual increase in the total number of students over the years, the proportion of diverse ethnicities relative to the total number of students per year remained more or less constant. This suggests that no significant change occurred in the ethnic composition of SAMS. Validation of these ethnicity inferences by comparison with real-world data about the ethnic composition of AMSP classes was not possible, as no such detailed information was available. Anecdotal evidence regarding the proportion of racial minority members in the pay grade range encompassing AMSP students (i.e., O4-O6) indicates that the gender inference (as seen in Figure 5) may overestimate minority proportions; for example, in 2017, the Department of Defense reported that O4-O6 ranks had a 25% minority composition (2017 demographic report, p. 29), increasing slightly in subsequent years (2018: 25.1%; 2019: 25.3%).

Using the STM algorithm with the year variable raised a methodological challenge. While the STM model can handle a temporal variable as a covariate, some functions of its R package may fail when a temporal variable is included. Following unsuccessful experimentation, we opted to treat the year variable as continuous, despite potential issues of independence (observations in a given unit of time may be dependent on those preceding it; Hannan and Tuma, 1979). However, since the students' choice of monograph topic is driven by individual curiosity or requests from Army headquarters, monographs are not influenced by those of preceding years. This justifies treating the year variable as continuous, which

⁸ Results were interpreted following the guidelines provided by McHugh (2012).

⁹ The authors are grateful to Dr. Andreas Beger for advice regarding some of these matters.

Table 1. *Comparison of predicted versus observed class totals*

Year	Predicted total (PreTotals)	Observed total (ObsTotals)
1985	23	23
1986	43	43
1987	47	47
1988	61	61
1989	39	39
1990	74	74
1991	89	89
1992	35	35
1993	78	78
1994	44	44
1995	58	58
1996	64	64
1997	55	55
1998	69	69
1999	52	52
2000	42	42
2001	57	57
2002	72	72
2003	86	86
2004	88	82
2005	81	78
2006	77	78
2007	88	82
2008	112	103
2011	146	146
2012	140	140
2014	160	160
2015	120	120

Note: Full overview of the timeframe under study and data sources is available in the online appendix on the OSF repository.

allowed us to analyze the influence of publication year on topic prevalence and trace changes in the ideational dimension of US Army PME.

The STM algorithm discerns latent dimensions in the corpus—the shared topics, or strings of words in their base form. On the one hand, this allows the researcher to survey the entire corpus using statistical learning methods; on the other hand, understanding the meaning of the topics requires subject matter expertise. While the STM R package provides detailed diagnostic tools for model selection and topic inspection¹⁰, unsupervised learning algorithms like STM lack definitive validation measures; hence, topic interpretation is largely based on expert judgment.

Due to space constraints, we focused on the change in prevalence over time of only the top-five topics in terms of their share across the corpus (table 2 provides an overview of the distribution of the top-ten topic shares). In STM, two key measures used to identify words with high influence and strong associations for each topic are: Highest Prob, which identifies the most frequent words in a topic, and Frequency of Exclusivity (FREX), which selects words that are both common and distinctive to a topic. FREX thus helps differentiate between topics by trading off between frequency and exclusivity, providing

¹⁰ An overview of steps taken in regard to these issues is included in the online appendix.

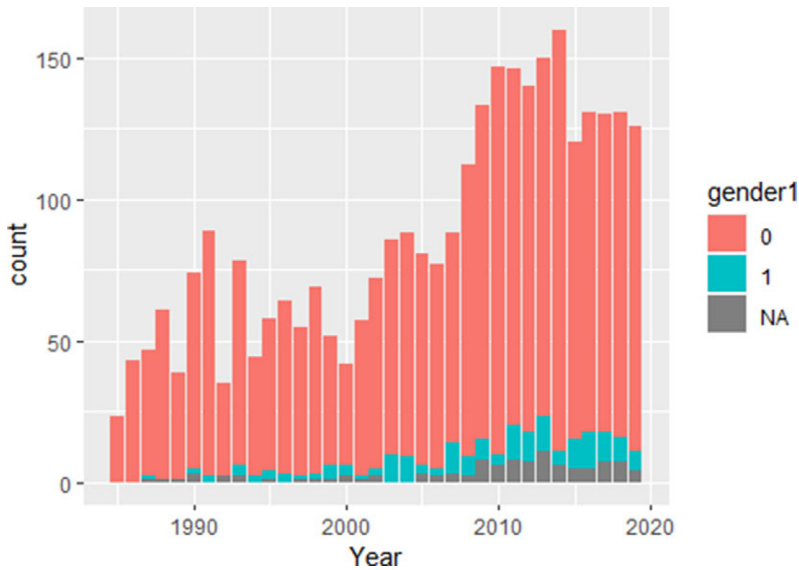


Figure 3. Distribution of inferred gender distribution per year ('gender1'), 1985–2019.

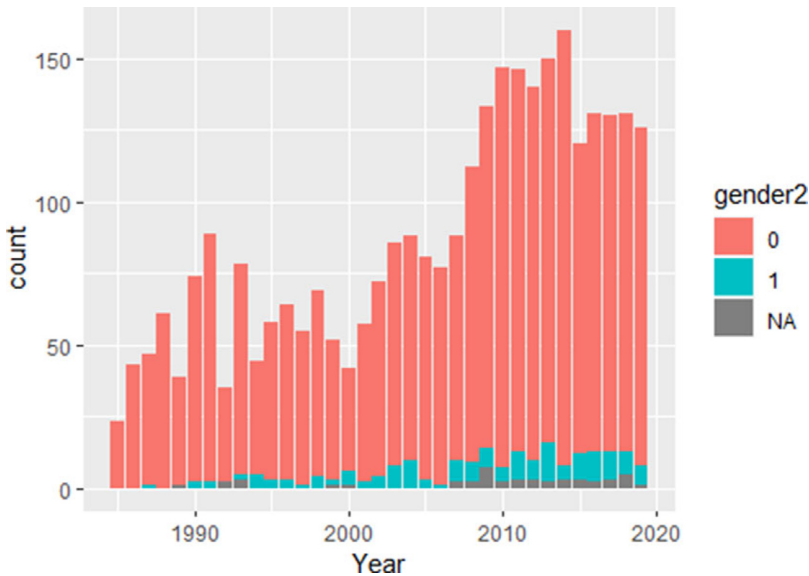


Figure 4. Distribution of inferred gender distribution per year ('gender2'), 1985–2019.

a comprehensive view of each topic. The results for the top five topics for the variable year are included in Table 3

Based on the Highest Prob and FREX results for each of the top five topics, the themes of these topics are presented in Table 4.

Topics 35 and 4 (joint military doctrine and strategy; and offensive and defensive operations) cover two tactical aspects of military operations. Keywords from Topic 35—such as doctrine, joint, campaign, and 3–0—indicate coordination among different military units and branches (e.g., armi, forc, militari), which prepares units for campaigns at the operational level (e.g., oper, command, object). Topic 4, based on words like enem, attack, manuev, defend, deep, and tempo, examines the operational art and tactics

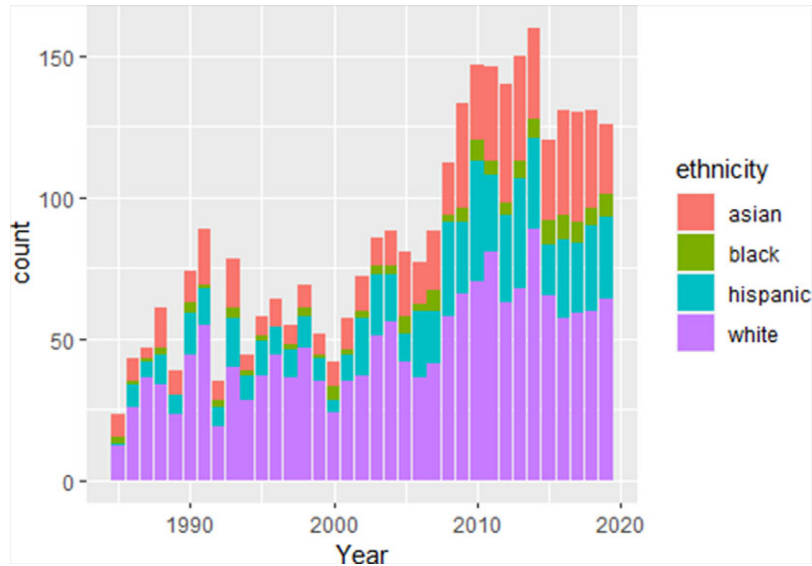


Figure 5. Distribution of inferred ethnicity among SAMS students, 1985–2019.

Table 2. Distribution of 10 top topic shares out of 50 topics, 1985–2019¹¹

Rank	Topic	Top words	Approx. share
1	35	oper, armi, doctrin	0.07
2	33	new, time, mani	0.065
3	4	forc, enemi, attack	0.06
4	31	use, effect, studi	0.058
5	6	forc, oper, unit	0.057
6	7	divis, armi, brigad	0.055
7	47	us, defens, strateg	0.052
8	9	state, nation, unit	0.05
9	39	cultur, state, polit	0.047
10	17	war, warfar, theori	0.045

involved in fighting an enemy and the methods and speed of maneuvering between offensive and defensive positions. Importantly, the words indicate that authors focused on large formations rather than small units.

Features of decision making specific to military operations form the core of Topics 33 and 31 (leadership and decision making; and risk analysis in military decision making). Topic 33 explores the human aspect of decision making (e.g., new, leader, said, go, just), focusing on challenges, experiences, and individual tactics. Topic 31, derived from analytical words such as risk, impact, model, result, and measure, investigates the structural metrics of evaluating risks and their impact on the decision-making process. Importantly, Topic 6 (full-spectrum training and education), characterized by words such as train, spectrum, deploy, OOTW (operations other than war), and SASO (stability and support operations),

¹¹ For clarity, Table 2 focuses on the top 10 most interpretable topics; the full 50-topic distribution is available in the online appendix in the paper’s Open Science Framework repository.

Table 3. Ten most probable and exclusive words per top five topics of variable ‘year’

	Topic 35		Topic 33		Topic 4		Topic 31		Topic 6	
	Highest Prob	FREX	Highest Prob	FREX	Highest Prob	FREX	Highest Prob	FREX	Highest Prob	FREX
Word 1	oper	oper	new	never	forc	enemi	use	risk	forc	deploy
Word 2	armi	object	time	believ	enemi	deep	effect	factor	oper	conting
Word 3	doctrin	3–0	mani	said	attack	depth	studi	impact	unit	train
Word 4	forc	campaign	leader	go	defens	maneuw	time	figur	armi	ootw
Word 5	militari	doctrin	general	put	oper	offens	chang	reduc	train	spectrum
Word 6	command	art	make	came	battl	defend	base	indic	mission	melt
Word 7	object	achiev	success	back	maneuw	tempo	result	model	deploy	forc
Word 8	joint	phase	take	hand	use	attack	risk	result	militari	mission
Word 9	campaign	strateg	year	just	tactic	surpris	analysi	tabl	capabl	saso
Word 10	war	fm	work	perhap	offens	penetr	requir	measur	task	task

Table 4. *Top five topics' themes*

Topic's number	Identified theme
35	Joint military doctrine and strategy
33	Leadership and decision making
4	Offensive and defensive operations
31	Risk analysis in military decision making
6	Full spectrum training and operations

encompasses broad-spectrum training and deployment initiatives. The words denote practices and measures implemented to prepare forces for a full range of missions.

Figure 6 provides an overview of the change in prevalence of these five topics between 1985 and 2019. The change is visualized by the direction (increase or decrease) and the degree (steepness) of each topic's trend line, with each line surrounded by its prediction interval.

Figure 6 shows a notable increase in the prevalence of Topic 35, joint military doctrine and strategy, which rose from about 3.5% to nearly 5% during the period examined. This supports the heightened focus that the US armed forces have placed on joint operations and a systemic doctrinal shift emphasizing joint operational frameworks. At the same time, a modest growth from just under 3% was noted in Topic 33 (leadership and decision making), reflecting SAMS' enduring focus on honing leadership abilities and the importance of effective leadership in fostering operational success.

In contrast, Topic 4 (offensive and defensive operations) underscores a dramatic decline in the topic representing large-scale campaigns, with the prevalence dropping from a significant 7% in 1985 to less than 1% in 2019. The shift likely indicates the transition of the American military after the Cold War; instead of large-scale engagements that predominated strategic thinking during the Cold War era, unconventional warfare started to loom larger on the strategic horizon, resulting in the reduced importance given to large-scale offensive and defensive operations (Lambeth, 2013; Romaniuk and Burgers, 2017).

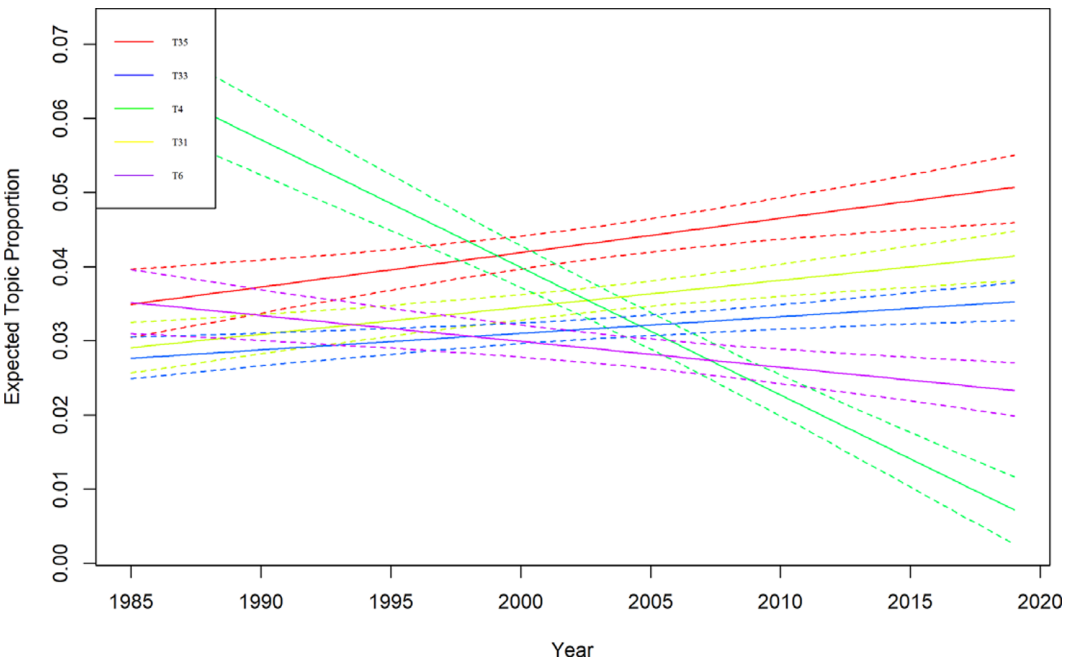


Figure 6. *Distribution of topic shares of 50 topics, 1985–2019.*

Initially hovering around 3%, Topic 31 (risk analysis in military decision making) registered a moderate increase to about 4%, underscoring the growing importance of formal risk assessment in high-stakes military decision-making processes (Bosley, 2017). Lastly, the prevalence of Topic 6 (full-spectrum training and education) began high but decreased over time, which mirrors the initial readiness efforts following the Cold War that examined a broad spectrum of scenarios. Over time, however, continued engagements of the Army in prolonged military campaigns, such as those in Iraq and Afghanistan, necessitated a more specialized operational readiness over broader, mission-oriented preparations.

5. Discussion

This paper presented an analytical strategy for overcoming conceptual and data access limitations associated with the study of PME institutions, which define officership as an institution. First, we developed a new conceptual framework of officership and PME as inter-related institutions embedded in the officer corps and PME organizations. Next, we used three types of machine-learning algorithms for gender inference, ethnicity inference, and topic modeling to trace changes in the US Army's PME institution, SAMS, from 1985 to 2019. The computational social science approach allowed us to extract new information from existing data; in this case, we reconstructed the gender and ethnic composition of consecutive SAMS classes and latent dimensions of their research corpus from authors' names and the full text of their monographs. We tested the viability of the conceptual framework and its operationalization using these computational tools to identify change (or lack thereof) across the time period.

Our study asked: "How have the demographic (gender and ethnicity) and doctrinal/ideational dimensions of SAMS changed from 1985 to 2019, and what do these changes reveal about institutional inertia and adaptation within the US Army?" Based on our theoretical framework, we hypothesized that: (1) due to institutional inertia, the demographic composition of SAMS would remain largely stable over time; and (2) critical junctures—such as the end of the Cold War and post-9/11 shifts—would drive significant changes in the doctrinal/ideational content, even as demographics remain unchanged.

The initial results suggest that no significant change took place in the gender and ethnicity dimensions of SAMS. The inter-coder reliability test indicated that the figures inferred by the gender R package, based on official US data, are more rigorous than those inferred by the GenderGuesser R package. Consequently, the gender composition, as approximated by the reliable gender1 variable, appears stable despite the overall increase in class size over the years. This finding supports Hypothesis 1, confirming that institutional inertia has maintained stable demographic patterns.

In contrast, the STM results indicate a meaningful change in the ideational dimension. The topic modeling analysis revealed shifts in thematic content over time, such as an increase in topics related to joint military doctrine and risk analysis, alongside a decline in topics associated with large-scale offensive and defensive operations. This observed shift in doctrinal content addresses Hypothesis 2, suggesting that external strategic pressures and critical junctures have driven changes in the intellectual and operational focus of SAMS, even though the underlying demographics remain constant.

We felt confident inferring the gender of research monographs' authors based on their first names. However, we were not confident inferring ethnicity from American last names, and the limited external data available precluded a robust statistical validation. Anecdotal evidence, such as the Department of Defense demographic reports for O4–O6 ranks, suggests discrepancies between inferred and real-world ethnic compositions, leading us to consider our ethnicity inferences as less reliable.

Using the STM algorithm with the year variable raised a methodological challenge. While the STM model can handle temporal variables as covariates, some functions of its R package failed when a temporal variable was included. Following unsuccessful experimentation, we opted to treat the year variable as continuous, which, despite potential independence issues, was justified by the fact that students' monograph topic choices are driven by individual curiosity or requests from Army headquarters. This allowed us to analyze the influence of publication year on topic prevalence and trace changes in the ideational dimension of US Army PME.

The study encountered limitations in validating the machine learning algorithms and required careful consideration of ethical and privacy issues when inferring gender and ethnicity from publicly available data. Furthermore, recent federal actions have removed key datasets in areas such as climate change, public health, and Diversity, Equity and Inclusion (DEI) initiatives, making direct demographic data increasingly inaccessible. This context underscores the critical role of indirect estimation methods—like those used in our study—for tracking institutional change.

The findings carry significant practical, theoretical, and policy implications. Computational social science tools can extract valuable insights from limited data, overcoming common military research constraints. Our new conceptual framework for analyzing officership and PME offers a novel lens for understanding military culture and the evolution of the officer corps, while also highlighting the need for robust privacy safeguards when using indirect methods for demographic inference. Moreover, the contrast between stable demographics and shifting doctrinal content indicates that while the structural composition of the officer corps remains unchanged, its intellectual orientation evolves in response to external pressures. This has important policy implications, particularly as direct demographic data become increasingly censored.

Future research should pursue cross-national comparisons and conduct deeper causal analyses to refine indirect estimation methods and better understand these dynamics. Specifically, researchers could compare the evolution of demographic and doctrinal dimensions across different military institutions in various national contexts, thereby testing the generalizability of our findings. In addition, employing advanced statistical techniques and longitudinal data designs would allow for a more precise identification of causal mechanisms underlying the stability of demographic patterns in the face of shifting strategic priorities. This deeper investigation may also inform the development of improved indirect estimation methods that can compensate for the growing inaccessibility of direct demographic data.

6. Conclusions

This study demonstrated and applied an analytical strategy to address both conceptual and data access challenges associated with the analysis of PME organizations as representations of the officer corps. It has shown that computational social science methods can extract valuable information from limited data sources, such as research monographs and their metadata, thereby overcoming traditional data constraints. The study also emphasized the importance of validating machine learning algorithms, which rely on access to real-world data, and highlighted the challenges that arise when such validation is limited by data availability.

The initial results imply that the gender and ethnicity dimensions of SAMS have remained largely unchanged over the period examined. In contrast, our structural topic model (STM) analysis revealed a significant shift in the ideational dimension—that is, in the subject matter taught, studied, and researched at SAMS—reflecting an adaptive transformation in military education. These results may suggest that the US Army—as an institution—has been more responsive to operational and strategic changes than to societal pressures, such as calls for increased diversity.

Additionally, the study raises important ethical considerations regarding the inference of gender and ethnicity from publicly available data, underscoring the need for careful privacy safeguards and ethical data practices. Overall, the findings carry several practical, theoretical, and ethical implications. They demonstrate the potential of indirect estimation methods, especially in an era of diminishing direct demographic data, and point toward the need for further research to explore these dynamics across different contexts.

Data availability statement. The replication code can be found on the Open Science Framework (OSF) website at <https://osf.io/sk4jm/>.

Acknowledgements. The authors would like to thank Colonel (Ret.) Dr. Kevin Benson and Dr. Andreas Beger for their support in the development of this article.

Author contribution. Conceptualization: T.L.; K.H. Methodology: T.L.; Data curation: T.L. Data visualisation: T.L.; K.H. Writing original draft: T.L.; K.H. All authors approved the final submitted draft.

Competing interests. The authors declare none.

References

- Amenta E and Ramsey KM** (2010) Institutional Theory. In K. T. Leicht & C. J. Jenkins (Eds.), *Handbook of Politics: state and society in global perspective* (pp. 15–39). New York: Springer. DOI: https://doi.org/10.1007/978-0-387-68930-2_2
- Benson KCM** (2010) *Educating the Army's Jedi: The School of Advanced Military Studies and the Introduction of Operational Art into U.S. Army Doctrine, 1983–1994* [Doctoral thesis, University of Kansas]. KU ScholarWorks. Available at <https://kuscholarworks.ku.edu/handle/1808/7716>.
- Blevins C and Mullen L** (2015) Jane, John... Leslie? A historical method for algorithmic gender prediction. *DHQ: Digital Humanities Quarterly* 9(3).
- Bosley LCC** (2017) The “new” normal: Instability risk assessment in an uncertainty-based strategic environment. *International Studies Review* 19(2), 206–227.
- Capoccia G and Kelemen RD** (2007) The study of critical junctures: Theory, narrative, and counterfactuals in historical institutionalism. *World Politics*, 59(3), 341–369. DOI: <https://doi.org/10.1017/S0043887100020852>
- Coccopuffs** (2020) *GenderGuesser: An R Package for Using the Genderize.io API to Guess the Gender of Names* [Computer software]. GitHub. Available at <https://github.com/coccopuffs/GenderGuesser>.
- Crosbie T, Lucas ER and Withande NE** (2019) Educating military elites: Professional military education in NATO countries. In Paananen S and Pulkka A-T (eds), *Processes and Practices in Military Training and Education*. Finnish National Defence University, pp. 41–60.
- Davis MS** (1980) *Development and Characteristics of National Defence Colleges—As a World Phenomenon* [Doctoral Dissertation]. Kingston: Queen's University.
- Ebbinghaus B** (2005) Can path dependence explain institutional change? Two approaches applied to welfare state reform (Discussion Paper 05/2). Retrieved from Max Planck Institute for the Study of Societies website https://www.mpifg.de/pu/mpifg_dp/dp05-2.pdf
- Government of Canada** (2019) *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans—TCPS 2* (2018). Available at https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2018.html.
- Government of Canada** (2022) *Open Data*. Available at <https://open.canada.ca/en/open-data>.
- Greener I** (2005) The potential of path dependence in political studies. *Politics*, 25(1), 62–72. DOI: <https://doi.org/10.1111/j.1467-9256.2005.00230.x>
- Grimmer J, Roberts ME and Stewart BM** (2022) *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.
- Hall PA and Taylor RCR** (1996) Political science and the three new institutionalisms. *Political Studies*, 44(5), 936–957. DOI: <https://doi.org/10.1111/j.1467-9248.1996.tb00343.x>
- Hannan MT and Tuma NB** (1979) Methods for temporal analysis. *Annual Review of Sociology* 5, 303–328.
- He Z and Schonlau M** (2021) Coding text answers to open-ended questions: Human coders and statistical learning algorithms make similar mistakes. *Methods, Data, Analyses* 15(1), 103–120.
- Holsting VS and Brænder M** (2020) A new spirit of militarism? Person–organization fit and economies of worth among new military cadets. In *Rethinking Military Professionalism for the Changing Armed Forces* (pp. 165–187). Springer.
- Immergut EM** (2006) Historical-institutionalism in political science and the problem of change. In A. Wimmer & R. Kössler (Eds.), *Understanding change: Methods, methodologies and metaphors* (pp. 237–59). London: Palgrave Macmillan. DOI: https://doi.org/10.1057/9780230524644_17
- Lambeth BS** (2013) Lessons from modern warfare: What the conflicts of the post–Cold War years should have taught us. *Strategic Studies Quarterly* 7(3), 28–72.
- Libel T** (2010) *Training and Education of the Combat Professional Officer Corps in the West during the Post-Modern Era* [Dissertation]. Bar Ilan University. [Hebrew].
- Libel T** (2016) *European Military Culture and Security Governance: Soldiers, Scholars and National Defence Universities*. Routledge.
- Libel T** (2020) The rise and fall of the study of the military profession: From the sociology of military profession to the sociology of security expertise. In Hachey KK, Libel T and Dean WH (eds.), *Rethinking Military Professionalism for the Changing Armed Forces* (pp. 12–28). Springer. Available at <https://link.springer.com/book/10.1007/978-3-030-45570-5>
- Libel T** (2021) Professional military education as an institution: A short (historical) institutionalist survey. *Scandinavian Journal of Military Studies* 4(1), 121–131. <http://doi.org/10.31374/sjms>
- McHugh ML** (2012) Interrater reliability: The kappa statistic. *Biochemia Medica* 22(3), 276–282.
- Murray W** (2000) Does military culture matter? *Orbis* 43 (1), 27–42
- Muth J** (2011) *Command culture: Officer education in the US army and the German armed forces, 1901–1940, and the consequences for World War II*. Denton: University of North Texas Press

- Roberts ME, Stewart BM and Airoidi E** (2016) A model of text for experimentation in the social sciences. *Journal of the American Statistical Association* 111(515), 988–1003. <https://doi.org/10.1080/01621459.2016.1141684>
- Romaniuk SN and Burgers TJ** (2017) Full spectrum dominance: US national security doctrine in the new global security environment. In Romaniuk SN and Grice F (eds), *Future of US Warfare*. Routledge, pp. 41–68.
- Segal DR** (2004) *Social and Cultural Dynamics of American Military Organizations*. United States Army Research Institute for the Behaviours and Social Sciences. Available at <https://apps.dtic.mil/sti/tr/pdf/ADA419659.pdf>.
- Song H, Tolochko P, Eberl JM, Eisele O, Greussing E, Heidenreich T and Boomgaarden HG** (2020) In validations we trust? The impact of imperfect human annotations as a gold standard on the quality of validation of automated content analysis. *Political Communication* 37(4), 550–572.
- Steinmo S** (2008) What is historical institutionalism? In D. Della Porta & M. Keating (Eds.), *Approaches and Methodologies in the Social Sciences: A Pluralist Perspective* (pp. 150–178). Newbury Park, CA: Sage.
- Van Creveld ML** (1990) *The Training of Officers: From Military Professionalism to Irrelevance*. Free Press.
- Welbers K Van Atteveldt W and Benoit K** (2017) Text analysis in R. *Communication methods and measures*, 11(4), 245–265.
- Xie F** (2021) *Predicting Ethnicity from Names with Rethnicity: Methodology and Application*. arXiv preprint. <https://doi.org/10.48550/arXiv.2109.09228>