

A RECIPE FOR DATA LITERACY: MODERNIZING
THE HUMAN DOMAIN

A thesis presented to the Faculty of the U.S. Army
Command and General Staff College in partial
fulfillment of the requirements for the
degree

MASTER OF MILITARY ART AND SCIENCE
General Studies

by

DONALD F. INGHAM, MAJOR, U.S. ARMY
Master of Professional Studies (Technology Management)
Georgetown University, Washington, DC, 2018

Fort Leavenworth, Kansas
2022

Approved for public release; distribution is unlimited. Fair use determination or copyright permission has been obtained for the inclusion of pictures, maps, graphics, and any other works incorporated into this manuscript. A work of the United States Government is not subject to copyright, however further publication or sale of copyrighted images is not permissible.

REPORT DOCUMENTATION PAGE				<i>Form Approved</i> <i>OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 10-06-2022		2. REPORT TYPE Master's Thesis		3. DATES COVERED (From - To) AUG 2021 – JUN 2022	
4. TITLE AND SUBTITLE A Recipe for Data Literacy: Modernizing the Human Domain				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Donald F. Ingham				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Command and General Staff College ATTN: ATZL-SWD-GD Fort Leavenworth, KS 66027-2301				8. PERFORMING ORG REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution is Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT What could the U.S. Army accomplish with an additional <i>11,341,330 annual work hours</i> ? As the Army undertakes a sweeping equipment modernization effort, the human domain will require a similar modernization effort to capitalize on the advantages afforded by new equipment and adequately prepare for Multi-Domain Operations. Many of these material upgrades rely on one fundamental resource – data. This research measures the correlation between data literacy training and the amount of time mid-grade U.S. Army Officers spend manipulating data (exporting online databases, changing excel spreadsheets, transferring data into PowerPoint, etc.) – instead of analyzing information. Evidence suggests that an increase of approximately 60 hours of Professional Military Education (PME) data literacy relevant training will reduce daily time requirements by 33 minutes. 33 minutes per-Officer, per day is a significant return on investment – not to mention the direct operational benefits of training next-generation thinkers who are able to truly operate in multiple warfighting domains.					
15. SUBJECT TERMS Data, Data Literacy, Processed, Informaion, Information Literacy, Education, Professional Military Education, PME, Basic Officer Leader Course, BOLC, Modernization					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT (U)	b. ABSTRACT (U)	c. THIS PAGE (U)			19b. PHONE NUMBER (include area code)
			(U)	129	

MASTER OF MILITARY ART AND SCIENCE
THESIS APPROVAL PAGE

Name of Candidate: Donald F. Ingham

Thesis Title: A Recipe for Data Literacy: Modernizing the Human Domain

Approved by:

_____, Thesis Committee Chair
LTC Christopher M. Baldwin, M.A.

_____, Member
Lynn M. Soots, Ph.D.

_____, Member
Eric M. Morrison, Ph.D.

Accepted this 10th day of June 2022 by:

_____, Assistant Dean of Academics for
Dale F. Spurlin, Ph.D. Degree Programs and Research

The opinions and conclusions expressed herein are those of the student author and do not necessarily represent the views of the U.S. Army Command and General Staff College or any other governmental agency. (References to this study should include the foregoing statement.)

ABSTRACT

A RECIPE FOR DATA LITERACY: MODERNIZING THE HUMAN DOMAIN, by Donald F. Ingham, 129 pages.

What could the U.S. Army accomplish with an additional *11,341,330 annual work hours*?

As the Army undertakes a sweeping equipment modernization effort, the human domain will require a similar modernization effort to capitalize on the advantages afforded by new equipment and adequately prepare for Multi-Domain Operations. Many of these material upgrades rely on one fundamental resource – data. This research measures the correlation between data literacy training and the amount of time mid-grade U.S. Army Officers spend manipulating data (exporting online databases, changing excel spreadsheets, transferring data into PowerPoint, etc.) – instead of analyzing information. Evidence suggests that an increase of approximately 60 hours of Professional Military Education (PME) data literacy relevant training will reduce daily time requirements by 33 minutes. 33 minutes per-Officer, per day is a significant return on investment – not to mention the direct operational benefits of training next-generation thinkers who are able to truly operate in multiple warfighting domains.

ACKNOWLEDGEMENTS

I want to thank my wife, Rachel Ingham, and son Daniel Ingham, for putting up with my obsession and constant absenteeism during this process. I could not have completed this work without your support. Rachel, your advice and dialogue has added significant depth to this analysis. Your ability to simply explain mathematical concepts is truly a gift. Additionally, I want to thank Dr. Lynn Soots for her bottomless well of excitement, encouragement, and assistance. I especially appreciate the significant amount of time you freely spent on nights, weekends, and holidays helping me through this process. I would not have completed this work without your support.

TABLE OF CONTENTS

	Page
ABSTRACT.....	iv
ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS.....	vi
ACRONYMS.....	ix
ILLUSTRATIONS.....	x
CHAPTER 1 INTRODUCTION.....	1
Data-bite.....	1
Background.....	6
Problem Statement.....	9
Purpose of the Study.....	10
Research Questions.....	10
Primary Research Question.....	11
Secondary Research Questions.....	11
Hypotheses.....	11
Assumptions.....	12
Scope.....	13
Limitations and Delimitations.....	13
Significance of the Study.....	14
Summary.....	15
CHAPTER 2 LITERATURE REVIEW.....	16
Introduction.....	16
Restated Research Questions.....	16
Primary Research Question.....	17
Secondary Research Questions.....	17
Review Analysis of Fundamental Doctrine: ADP 6-0 Mission Command.....	18
Review Analysis of Tactical Doctrine: FM 6-0 Commander and Staff Organization and Operations.....	21
Review Analysis of Signal Doctrine: FM 6-0.2 Signal Support to Operations.....	22
Cognitive Processes: Cognitive Techniques in Visual Data Interpretation.....	23
Civilian Equivalence: How Many Ways Can We Teach Data Literacy?.....	25
Current State (Civilian): Lead with Data.....	27
Deep Dive: Be Data Literate.....	28
Civilian Training: Teaching A One-Credit Course on Data Literacy and Data Visualization.....	31

Summary	32
CHAPTER 3 RESEARCH METHODOLOGY	35
Introduction.....	35
Restated Research Questions	35
Primary Research Question.....	35
Secondary Research Questions	36
Method	36
Data Collection	40
CGSOC Class of 2022 Survey	40
Determining the Population and Sample	41
Determine the Type of Survey	41
Designing the Survey Questions	42
Receive CGSOC Institutional Review Board (IRB) Approval.....	42
Distribute the Survey	42
Safeguard all Survey Materials.....	42
Analyze the Clean Data	43
DA PAM 600-3 Smartbooks.....	44
Analyze the Clean Data	45
Basic Officer Leadership Course (BOLC) Programs of Instruction (POIs)	45
Analyze the Clean Data	48
Data Analysis Process.....	48
Statistical Analysis.....	49
Ethical Considerations	53
Summary.....	56
CHAPTER 4 Data Findings and Analysis	58
Introduction.....	58
Restated Research Questions	58
Primary Research Question.....	58
Secondary Research Questions	58
Data Analysis.....	59
Exploratory Analysis Findings	59
CGSOC Class of 2022 Survey	59
DA PAM 600-3.....	63
Basic Officer Leadership Course (BOLC) Programs of Instruction (POI).....	64
Primary Research Question Findings.....	69
Secondary Research Question 1 Findings.....	73
Secondary Research Question 2 Findings.....	75
Secondary Research Question 3 Findings.....	77
Secondary Research Question 4 Findings.....	78
Summary.....	80
CHAPTER 5 Conclusions and Recommendations	84

Introduction.....	84
Implications	86
Three Recommendations	87
Recommendations for Future Research	88
Conclusion	89
GLOSSARY	92
APPENDIX A: Survey Questions and Intent	95
APPENDIX B: Survey Recruitment Email	99
APPENDIX C: Survey Informed Consent.....	100
APPENDIX D: POI Document Analysis.....	104
APPENDIX E: Permission Memos	110
APPENDIX F: DA PAM 600-3 Smartbook Document Analysis.....	115
APPENDIX G: Power BI Measures	116
APPENDIX H: Army Branch Acronyms	117
Bibliography	118

ACRONYMS

AI	Artificial Intelligence
BOLC	Basic Officer Leadership Course
CGSOC	Command and General Staff Officers College
COP	Common Operational Picture
DA PAM	Department of the Army Pamphlet
DL	Data Literacy
DTMS	Digital Training Management System
ESR	Equipment Status Report
GARMY	Global Combat Support System - Army
HMMWV	High Mobility Multipurpose Wheeled Vehicle
ICH	Instructor Contact Hours
IRB	Internal Review Board
LMTV	Light Medium Tactical Vehicle
MDO	Multi-Domain Operations
ML	Machine Learning
NEC	Network Enterprise Center
OE	Operating Environment
PL	Platoon Leader
PME	Professional Military Education
POI	Program of Instruction
SPSS	Statistical Product and Service Solutions
TDC	Training Development Capability
XO	Executive Officer

ILLUSTRATIONS

Figure 1. Marginal Product/Processed Data.....	3
Figure 2. Useful Product/Visualized Information	4
Figure 3. Food Comparison to Data and Information	6
Figure 4. Current State	7
Figure 5. Desired State	8
Figure 6. Scope.....	13
Figure 7. The Commander’s Role in the Operations Process	18
Figure 8. Achieving Understanding	19
Figure 9. The IME-I Cognitive Information Model.....	24
Figure 10. Methodology Logic Chart.....	37
Figure 11. Data Model.....	39
Figure 12. Base Visualization	44
Figure 13. Conversion Issues	46
Figure 14. Power Query Editor Data Normalization.....	47
Figure 15. Data Analysis Measures.....	49
Figure 16. Example Primary Research Question Data Visualization (not actual data)	51
Figure 17. Example Secondary Research Question Data Visualization (not actual data)	52
Figure 18. Mid-grade U.S. Army Perspective on Data Literacy Training in PME.....	60
Figure 19. Average Total Time Comparison.....	61
Figure 20. BOLC Impact on Time – Overall	61
Figure 21. BOLC Impact on Perception of Excel Familiarity	62
Figure 22. BOLC Impact on Time – Comparison.....	63
Figure 23. DA PAM 600-3 Data Importance	64

Figure 24. POI Quantitative Analysis	65
Figure 25. Initial POI Qualitative Analysis.....	66
Figure 26. POI Qualitative Analysis – Branch Comparison	67
Figure 27. POI Qualitative Analysis – Removing Cyber.....	68
Figure 28. SPSS Results – Hypotheses H1-H3	71
Figure 29. SPSS Results – Hypotheses H4-H6	73
Figure 30. Data Literacy Training Reduces Manipulation Time	74
Figure 31. Data Literacy Training Might Reduce Analysis Time.....	76
Figure 32. Data Literacy Training Might Reduce Perceived Excel Familiarity	78
Figure 33. No Evidence to Support Correlation between Data Literacy Training and Perception	79
Figure 34. Potential Impact of Data Literacy Training on Time.....	86
Figure 35. Recommended Data Literacy Learning Objectives	88
Figure 36. Data Literacy Class Final Project/Visualized Information	90

CHAPTER 1

INTRODUCTION

Data-bite

Imagine for a moment that mission success depends upon an Army staff's ability to cook. Each Staff/General (S/G) shop is responsible for a different course: the G1 (human resources) handles the soup, the G3 (operations) focuses on the main course, the G8 (finance) creates the desert, etc. The Commander's wartime performance is then dependent on the overall quality of his or her meal. Assuming that all cooking equipment and raw ingredients are equally distributed, each section would inevitably provide a course to the Commander; with wildly inconsistent results. As cooking is (rightfully) not included in Army Officer Professional Military Education (PME), the individual initiative to learn this skill is the only significant remaining performance variable. The resulting difference in quality is largely dependent upon how much cooking experience individuals have in their personal life. If the G1 provides a delicious French onion soup, but the G3 assumes a dry tuna sandwich will suffice for the main course, would the Commander ultimately succeed?

As the Army continues to modernize, our situation is increasingly looking like a cooking competition, with data being the prize ingredient. This organization creates a colossal amount of data (raw ingredients) and provides basic analytical tools (cooking equipment) in the form of Microsoft Excel, and Power BI, but largely relies on individual initiative to learn the fundamental skills required to correctly use these inputs to provide high-quality products (courses/meals).

A fundamental task of staff work, regardless of section, is turning data into information and eventually knowledge/understanding.¹ This truism is broadly, and repeatedly, articulated in doctrine, and yet is largely absent from standard civilian or military education programs. This research has determined that, excluding Cyber, less than one% of the Basic Officer Leadership Course (BOLC) institutional training time relates to data literacy, and nine of the seventeen branches have zero data literacy relevant instruction.

All meals are not created equal. Similarly, all staff products are not equal. An organization can subsist on marginal products for a time but will find it difficult to exploit advantages, predict requirements, and take advantage of opportunities. For example, in 2019, the 35th Theater Tactical Signal Brigade (TTSB) used Figure 1 as their Status of Funds (SOF) to manage an annual budget of over \$10,000,000. This weekly product required between two and four hours of formatting. The finished document provided only minimal value to select individuals. Just because data has been “processed” does not mean that the resulting product depicts relevant information or supports decision making.

Fund	Funds Center	Cost Collector	Commitment item	Allotment	Comm. (Cum)	Oblig. (Cum)	Undisbursed Obligations	Open Expenses	Disb. (Cum)
HHC 35th SIG BDE									
202010D20	A76CC		21T0		38,735.61	38,735.61	37,681.27	2,710.79	1,054.34
202010D20	A76CC	C/0040167565	21S0		3,250.00	3,250.00	3,250.00	0.00	0.00
202010D20	A76CC		262Q		222.60	222.60	222.60	0.00	0.00
				0	42,208.21	42,208.21	41,153.87	2,710.79	1,054.34
67th ESB									
202010D20	A76CC	C/0040181158	21T0		4,447.08	4,447.08	4,118.18	328.90	328.90
202010D20	A76CC		260G		2,790.06	2,790.06	2,790.06	0.00	0.00
202010D20	A76CC		2615		103,014.75	103,014.75	103,014.75	0.00	0.00
202010D20	A76CC		260J		31.30	31.30	31.30	0.00	0.00
202010D20	A76CC		261Q		16,223.00	16,223.00	16,223.00	0.00	0.00
202010D20	A76CC		2625		21,366.78	21,366.78	21,366.78	0.00	0.00
202010D20	A76CC		2615		376.46	376.46	376.46	0.00	0.00
202010D20	A76CC	C/0040181159	2615		393.67	393.67	1,097.67	393.67	-704.00
202010D20	A76CC	C/0040181186	2625		238.27	238.27	238.27	88.36	0.00
202010D20	A76CC		261A		5,397.47	5,397.47	5,397.47	716.00	0.00
202010D20	A76CC		262A		290.26	290.26	290.26	40.72	0.00
202010D20	A76CC		262Q		724.26	724.26	724.26	514.92	0.00
202010D20	A76CC		2615		2,195.34	2,195.34	2,195.34	2,195.34	0.00
202010D20	A76CC	C/0040181213	2625		956.15	956.15	956.15	10.09	0.00
202010D20	A76CC		260B		26,164.13	25,927.77	25,927.77	0.00	0.00
202010D20	A76CC		261A		-3,732.17	-3,732.17	691.83	-4,381.46	-4,424.00
202010D20	A76CC		262A		4,880.42	4,880.42	4,880.42	192.08	0.00
202010D20	A76CC		262Q		446.28	446.28	446.28	356.72	0.00
202010D20	A76CC		263A		0.04	0.04	0.04	0.00	0.00
202010D20	A76CC		310A		1,984.89	1,982.87	1,982.87	0.00	0.00
202010D20	A76CC		2615		1,463.56	1,463.56	1,463.56	0.00	0.00
202010D20	A76CC	C/0040181241	260B		2,693.80	2,693.80	2,693.80	0.00	0.00
202010D20	A76CC		262A		3,150.67	3,150.67	3,150.67	0.00	0.00
202010D20	A76CC		262Q		70.84	70.84	70.84	0.00	0.00
202010D20	A76CC		310A		354.30	134.80	134.80	0.00	0.00
				0	195921.61	195463.73	200262.83	455.34	-4799.10
51st ESB									
202010D20	A76CC	C/0040181256	21T0		31,073.01	31,073.01	31,073.01	0.00	0.00

Figure 1. Marginal Product/Processed Data

Source: Created by author.

To correct this issue, the Command created Figure 2. This dashboard displays the same data, yet it only takes five minutes to update, while better communicating impacts. This product significantly expedited decision-making, transparency, and efficiency, which directly resulted in saving over \$2,000,000 in funding.

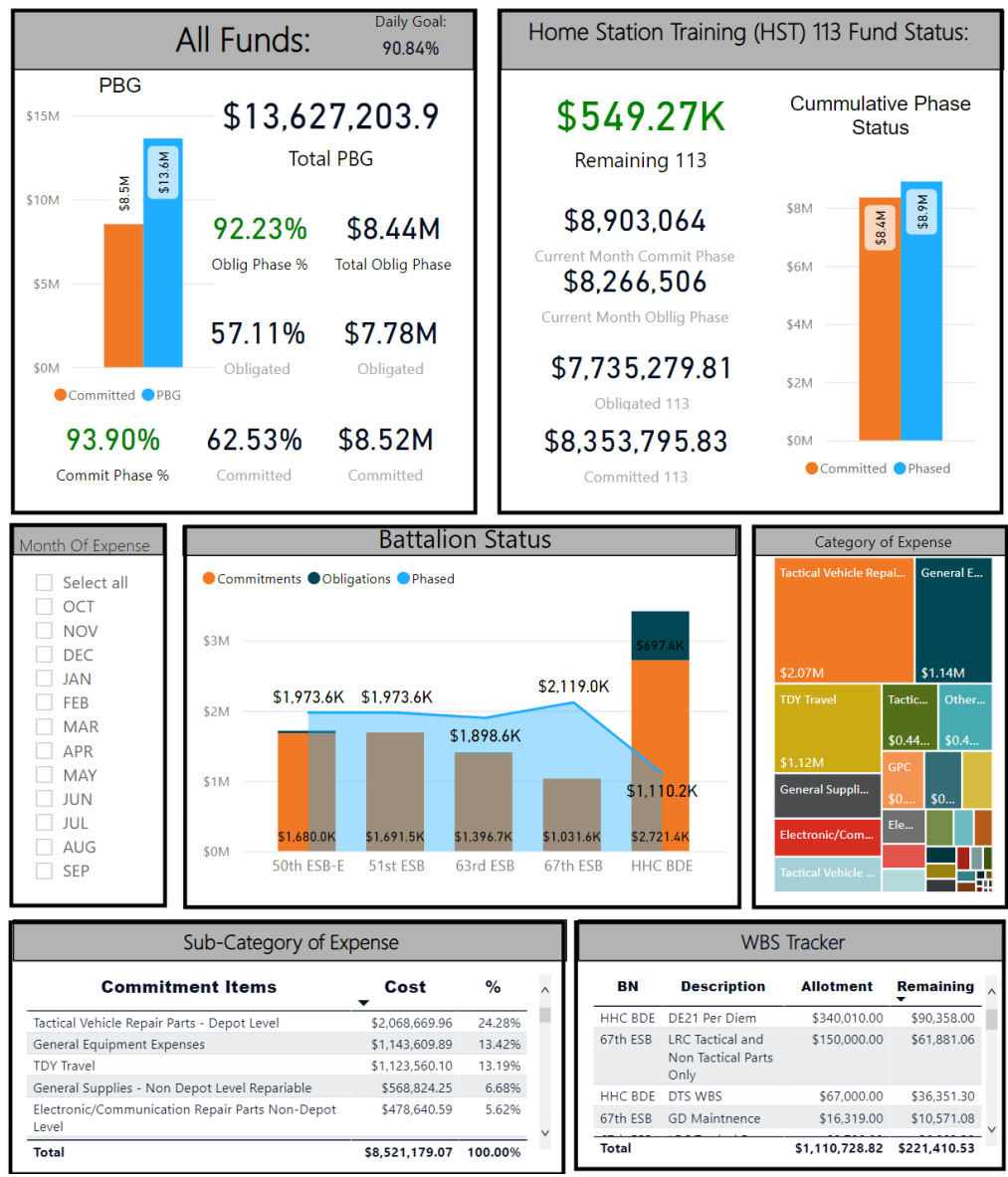


Figure 2. Useful Product/Visualized Information

Source: Created by author.

The creation of Figure 2 would have been impossible if decision-makers did not possess a working knowledge of data management. Beyond this basic understanding of data, nothing was unique about the 35th TTTSB. The required software is free, created by Microsoft, and available on government networks. As the quantity of U.S. Army data grows, a fundamental understanding of this resource becomes increasingly valuable. Figure 3 helps emphasize this point. Processed data may sustain an organization in the same way raw peanut butter can sustain life, but it lacks relevancy at most tables. Until Army staffs learn to use all the raw ingredients at their disposal, it will be difficult to elevate products beyond mere descriptive analysis. A basic understanding of data enables formations at every level of warfare to predict requirements, exploit modernization initiatives, and become truly data-centric.

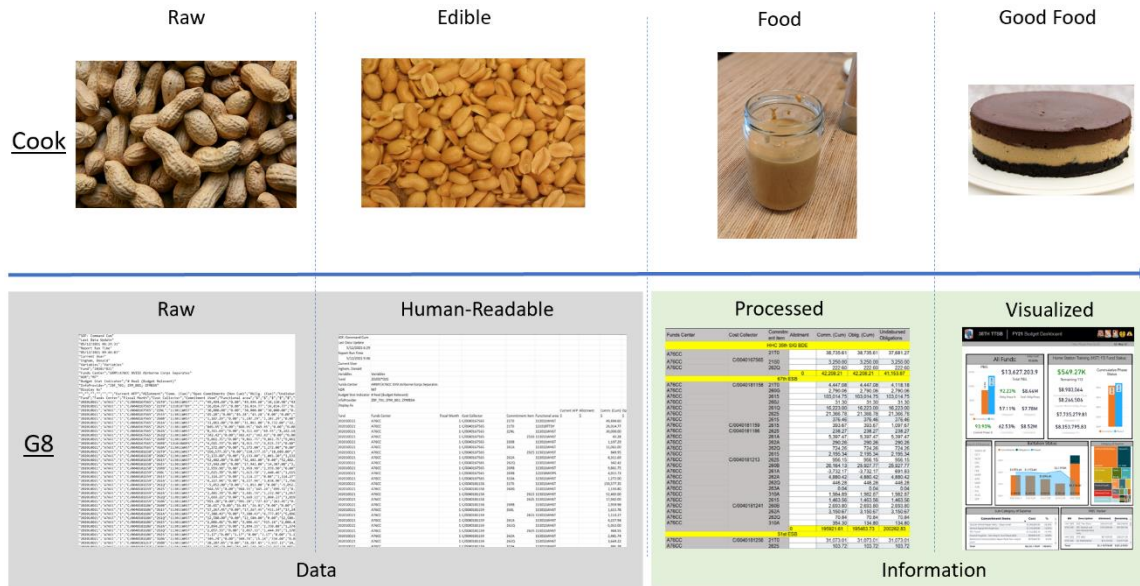


Figure 3. Food Comparison to Data and Information

Source: Created by author using images from Google, “Peanut,” Creative Commons, accessed 22 April 2022, <https://p1.pxfuel.com/preview/1004/417/215/peanuts-food.jpg>; Google, “Peanut,” Creative Commons, accessed 22 April 2022, <https://p1.pxfuel.com/preview/72/962/424/peanuts-shell-nuts-delicious.jpg>; Google, “Peanut Butter,” Creative Commons, accessed 22 April 2022, https://upload.wikimedia.org/wikipedia/commons/thumb/b/b0/Peanut_butter_glass_3.jpg/900px-Peanut_butter_glass_3.jpg?20210227133103; Google, “Peanut Butter Dessert,” Creative Commons, accessed 22 April 2022, https://live.staticflickr.com/4111/5207974469_e2dcbaeca2_b.jpg.

Background

The Army faces a rapidly changing Operating Environment (OE), with data becoming fundamental to increasingly complex operations. Multi-Domain Operations (MDO) necessitates enhanced service integration and interoperability, which is technologically enabled through initiatives such as Project Convergence.² Unfortunately, while these technical problems are addressed, data remains siloed within stove-piped systems that require highly specific and non-transferable training to operate individual interfaces. For example, the Property Book Unit Supply Enhanced (PBUSE) transition to

the Global Combat Support System – Army (GARMY) required extensive systems-based retraining (both formal and on-the-job) to ensure basic proficiency in the new layout, tools, and features.

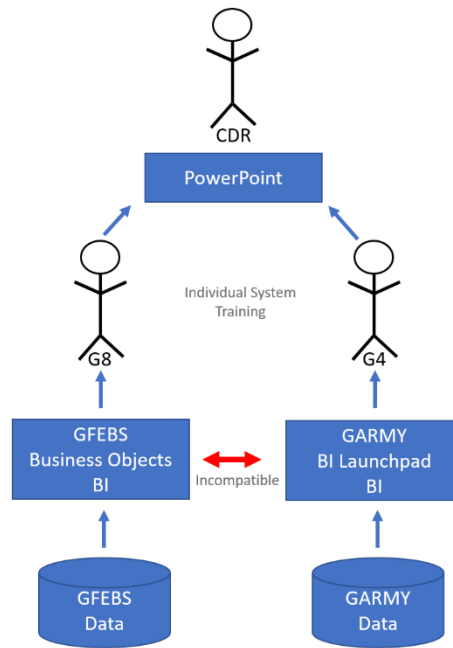


Figure 4. Current State

Source: Created by author.

The Army currently prefers to train Soldiers on individual systems rather than educate leaders on what to do with the underlying data. Logistics (G4) Soldiers learn to use Global Combat Support System – Army (GCSS Army or GARMY), while the Finance (G8) Corps uses the General Fund Enterprise Business System (GFEBs), and the Operations section uses Digital Training Management System (DTMS) (Figure 4). These stovepipes require leaders to have basic proficiency with multiple systems, frequent touchpoints with each, and necessitate manual data consolidation techniques that waste

our most valuable resource – time. Instead, if the entire Officer Corps is able to integrate any available relevant data, significant time can be better allocated towards information analysis rather than information creation (Figure 5).

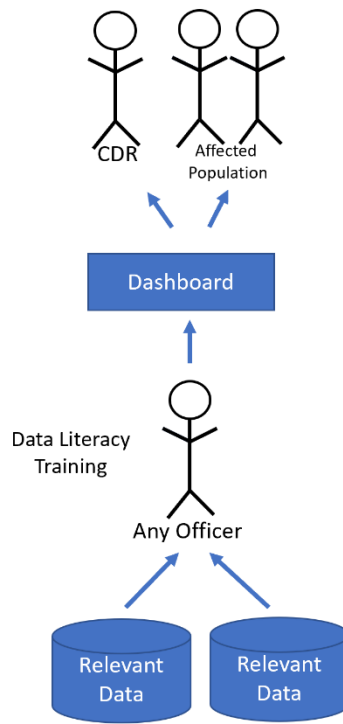


Figure 5. Desired State

Source: Created by author.

As technology continues to evolve, it is increasingly integrated into the decision-making cycle. A conceivable near-term reality is one where technology begins to recommend decisions that end lives. When this happens, is it truly ethical to blindly accept the output of these systems without first understanding the input? Incomplete,

manipulated, or dirty (incorrect or biased) data will have catastrophic and lethal consequences if humans cannot recognize and correct these issues.

Soldiers have a finite number of work hours, yet reporting requirements continue to grow. With the correct training, flexible unit-level process automation is entirely possible. Correctly implemented, automation will return time to our formations and enable our people to conduct higher-level tasks, such as analysis. Inevitably, this re-prioritization will result in better (and faster) decisions. Once current initiatives succeed in standardizing data between services, an educated and experienced force will be able to exploit these advancements to enable success on the multi-domain battlefield.

Problem Statement

Currently, U.S. Army Officers are not trained to process data to the degree of efficiency and effectiveness that our doctrine requires. Higher-level data processing is a required part of an Officer's job; yet, it is neither a mandatory undergraduate requirement nor directly addressed in most Army Professional Military Education (PME). Doctrine relies heavily on the assumption that Soldiers intrinsically understand how to transform data into information and knowledge; yet, the training for the affected population is not provided. Variations in branch and/or institutional training may contribute to the current differences in Officers' understanding of data. This discrepancy likely results in the Army wasting a significant amount of time manipulating data, rather than analyzing information.

Purpose of the Study

The military is in the process of making significant investments in Machine Learning (ML), Artificial Intelligence (AI), and data processing.³ During this modernization effort, there is value in exploring the force-wide understanding of the resource that makes these wonderful and dangerous innovations possible. This study seeks to investigate the amount of time allotted in formal Officer training programs and the post-training opinions of mid-grade Officers as a requirement for their jobs. The research aims to establish a baseline of how well our current force understands and uses data to determine if the Army collectively possesses an adequate level of comprehension to support operations efficiently and effectively through a sweeping modernization effort.

To meet the purpose of the study, an informational data review investigated the amount of time the U.S. Army allocates for data-relevant coursework within the Basic Officer Leadership Course (BOLC). This data aligns with data elements from a survey of mid-grade U.S Army Officers attending resident CGSOC. Combined, they address a correlation between perceived skill level and time allotted for training data processing.

Research Questions

In order to define the current state of the U.S. Army's understanding of data, two categories were considered: the input and the output. The input consists of two elements: characteristics required of officers (drawn from Department of the Army Pamphlet 600-3 Smartbooks), and institutional training (drawn from a review of the Basic Officer Leader Course Programs of Instruction). The output consists of a survey, targeting the Command and General Staff Officer College (CGSOC) population (addressing elements such as respondents' time spent manipulating data). As the CGSOC student population holds the

rank of either senior Captain or junior Major, this survey will specifically address the “mid-grade” Officer population. This analysis will be conducted to answer the primary research question of:

Primary Research Question

Is there a significant difference in the average amount of time mid-grade U.S. Army Officers spend manipulating data or analyzing information comparing across BOLC attended or current branch?

Four secondary research questions will be explored in order to provide clarity and depth to this analysis:

Secondary Research Questions

1. How much does Professional Military Education data literacy training change the amount of time mid-grade U.S. Army Officers spend manipulating data?
2. How much does Professional Military Education data literacy training change the amount of time mid-grade U.S. Army Officers spend analyzing information?
3. How much does Professional Military Education data literacy training change a mid-grade U.S. Army Officer’s familiarity with Excel?
4. How much does Professional Military Education data literacy training change a mid-grade U.S. Army Officer’s perception of their own data literacy?

Hypotheses

The first set of hypotheses follow:

H1: The average amount of time a mid-grade Officer spends manipulating data or analyzing information is different when comparing PME BOLC attended.

H2: The average amount of time a mid-grade Officer spends manipulating data is different when comparing PME BOLC attended.

H3: The average amount of time a mid-grade Officer spends analyzing information is different when comparing PME BOLC attended.

H0: The average amount of time a mid-grade Officer spends manipulating data or analyzing information is the same regardless of PME BOLC attended.

The second set of hypotheses follow:

H4: The average amount of time a mid-grade Officer spends manipulating data or analyzing information is different when comparing their current branch.

H5: The average amount of time a mid-grade Officer spends manipulating data is different when comparing their current branch.

H6: The average amount of time a mid-grade Officer spends analyzing information is different when comparing their current branch.

H0: The average amount of time a mid-grade Officer spends manipulating data or analyzing information is the same regardless of current branch.

Assumptions

The first assumption this research makes is that the Officer population possesses the ability to be trained to better understand data. The second assumption is that survey participants will answer honestly and accurately. The third assumption is that survey participants will answer questions based on the definitions of the terms given in the survey. Finally, this research assumes that a Soldier who is data literate is able to complete data-oriented tasks, whereas a person who is not data literate is more likely to fail at completing the same task.

Scope

The scope of this research is limited to the cognitive processes Soldiers use to turn data into information (current state versus desired state) (Figure 6). The scope of this research includes a comparison of how the Army views the data processing issue, and how the civilian education sector has approached the problem of data literacy training in relevant civilian research.

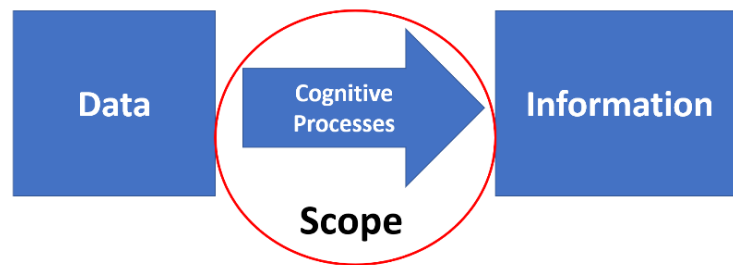


Figure 6. Scope

Source: Created by author.

Limitations and Delimitations

The primary limitation of this work is the amount of available time. This is influenced by competing academic requirements during CGSOC. As initially designed, the study explored additional areas such as how data presentation affects decision making, but at the recommendation of the committee, the research was scoped to focus on the foundational issue of data literacy. Given time, the researcher would have additionally increased the dataset to include the Captains Career Course contact hours.

A delimitation of this work is the CGSOC environment. Data collection was restricted to the constraints of the CGSOC population: mid-grade CGSOC Officers. Mid-grade is defined as holding the ranks of senior Captain to junior Major. Fortunately, this is a good population, given the purpose of the study. Additionally, this research settled on the term Data Literacy, and a singular definition, for further exploration despite finding numerous different educational categories that may address the Army's doctrinal requirement to understand how to use data. Finally, this research assumes that the civilian population is reflective of the pre-commissioning population. This is necessitated due to a lack of research that specifically addresses the understanding of data amongst pre-commissioning Cadets.

Significance of the Study

As the Army begins a period of rapid modernization with the advent of Multi-Domain Operations (MDO) and the prioritization of joint inter-operability, second and third order effects will begin to manifest in unforeseeable ways. Recent modernization efforts rely heavily on data to enable the wonderful and powerful advances they promise. This study tries to understand the current force-wide understanding of data as a resource that can be manipulated and exploited to drive the Commander's visualization process. As data underlies many of these innovations, this study investigated if the current force is capable of operating in a future environment saturated with data-heavy technologies such as Artificial Intelligence (AI) and Machine Learning (ML). Modernizing equipment without also appropriately training personnel to function in the new environment can severely limit any advantages new equipment might offer.

Finally, this study identified civilian best practices and recommend ways of implementing high-potential techniques within a military context. As the hypothesis of the study indicates significance, an escalation of focus and resources into education and training has the potential to significantly increase efficiency, speed, and accuracy at every level of warfare. Even the results that did not indicate a correlation hold value in providing a starting point for continued analysis or follow-on studies.

Summary

Similar to food, data is a perishable resource. Data can provide critical insights and drive operations in an efficient and effective way, but only if trained professionals know how to exploit this resource. There is a fundamental gap between the doctrinal expectation that Officers innately know how to turn data into information and the reality of the dearth of data-oriented content taught in PME. This problem is not unique to the military. The civilian sector provides examples of how to train a workforce to operate in a modern and interconnected environment. This gap, its implications, and high-potential civilian best practices will be explored further in the following section.

¹ Headquarters, Department of the Army (HQDA), *Army Doctrine Publication (ADP) 6-0, Mission Command* (Washington, DC: Army Publishing Directorate, 2019).

² Headquarters, Department of the Army (HQDA), *Army Multi-Domain Transformation: Ready to Win in Competition and Conflict*, Chief of Staff Paper #1, unclassified version (Washington, DC: Army Publishing Directorate, March 16, 2021, <https://api.army.mil/e2/c/downloads/2021/03/23/eeac3d01/20210319-csa-paper-1-signed-print-version.pdf>).

³ Ibid.

CHAPTER 2

LITERATURE REVIEW

Introduction

In order to address a wicked problem, doctrine begins by framing the operational environment. In regards to the Army's understanding of Data, it is important to define both the current state as well as the desired end state.¹ There is a distinct contrast between the current civilian state and the current state within the Army. Literature in this chapter address the current understanding of data in the Army and the education (Army and civilian) that underpins this current understanding, before merging best practices from both military and civilian institutions to outline a potential desired end state.

Restated Research Questions

Two categories were considered: the input and the output. The input consists of two elements: characteristics required of officers (drawn from Department of the Army Pamphlet 600-3 Smartbooks) and institutional training (drawn from a review of the Basic Officer Leader Course Programs of Instruction). The output consists of a survey, targeting the Command and General Staff Officer College (CGSOC) population, which addresses elements such as respondents' perception of their individual understanding of data. As the CGSOC student population holds the rank of either senior Captain or junior Major, this survey will specifically address the "mid-grade" Officer population. This analysis will be conducted to answer the primary research question of:

Primary Research Question

Is there a significant difference in the average amount of time mid-grade U.S. Army Officer's spend manipulating data or analyzing information comparing across BOLC attended or current branch?

Four secondary research questions will be explored in order to provide clarity and depth to this analysis:

Secondary Research Questions

1. How much does Professional Military Education data literacy training change the amount of time mid-grade U.S. Army Officers spend manipulating data?
2. How much does Professional Military Education data literacy training change the amount of time mid-grade U.S. Army Officers spend analyzing information?
3. How much does Professional Military Education data literacy training change a mid-grade U.S. Army Officer's familiarity with Excel?
4. How much does Professional Military Education data literacy training change a mid-grade U.S. Army Officer's perception of their own data literacy?

Review Analysis of Fundamental Doctrine: ADP 6-0 Mission Command

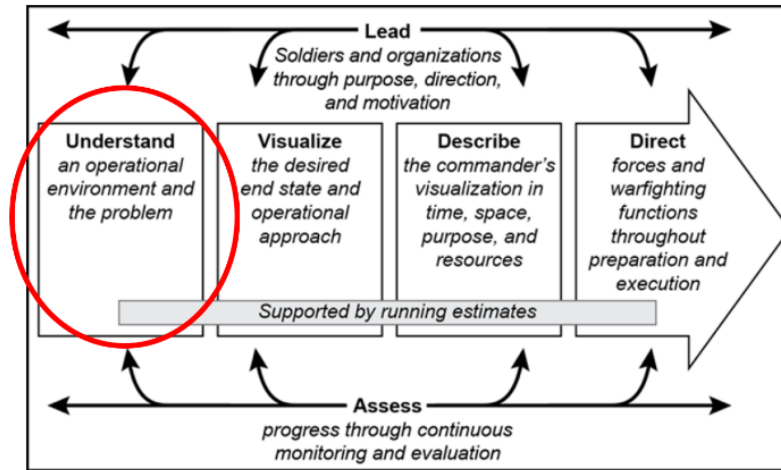


Figure 7. The Commander's Role in the Operations Process

Source: Headquarters, Department of the Army, Army Doctrine Publication 6-0, *Mission Command* (Washington, DC: Army Publishing Directorate, 2019), 2-14.

Commanders in the Army have final decision-making authorities. Commanders exercise this authority through their role in the operations process, which requires Commanders to first understand operational environments so decisions can be made within context (Figure 7).² It is the staff's job to help Commanders quickly and accurately reach this understanding. To do this, doctrine states that the first three steps begin with data, which is processed into information (Figure 8).³

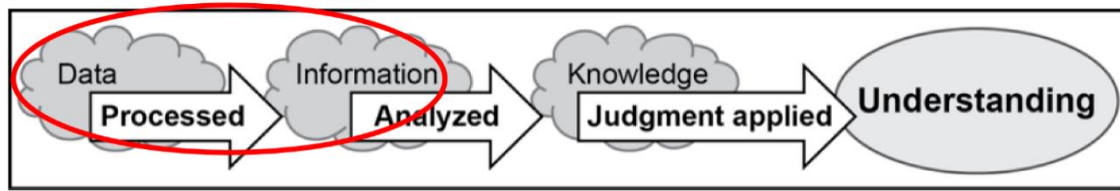


Figure 8. Achieving Understanding

Source: Headquarters, Department of the Army, Army Doctrine Publication 6-0, *Mission Command* (Washington, DC: Army Publishing Directorate, 2019), 2-4.

Words have meaning, so it is only fair to begin with the doctrinal definitions of the aforementioned terms. The scope of this paper only includes these first three steps thus the final steps will not be included in this review. “In the context of decision making, data consists of unprocessed observations detected by a collector of any kind (human, mechanical, or electronic).”⁴ “In the context of decision making, information is data that has been organized and processed in order to provide context for further analysis.”⁵

To clarify the definitions, the Army places each definition within the context of decision-making, because decisions are the desired output from the interpretation of information. The objective of information creation is not a purely academic endeavor; it is created to enable educated decisions. Within the confines of the definitions, there is one important term that is noticeably absent: *processed*. For context, in this publication (ADP 6-0), the term data is referenced 42 times, whereas the term *processed* is only used three times, and never defined on its own.⁶ This term should provide readers with the answer to the question of how data turns into information, but it is left undefined within ADP 6-0.

Instead, ADP 6-0 presents the term information management, which is not fully developed to the point of a clear and concise definition. Upon first reading, the term information management reflects the idea that it specifically refers to already processed data. ADP 6-0 defines information management as “the science of using procedures and information systems to collect, process, store, display, disseminate, and protect data, information, and knowledge products.”⁷ This definition reveals that the Army uses this term to include the management of data. ADP 6-0 includes a reference to FM 6-0.2⁸ for further clarification of this term, which will be discussed in the section entitled Review Analysis of Signal Doctrine: FM 6-0.2 below.

ADP 6-0 describes Information Services and Data as a way to “facilitate timely and accurate decision making and execution by processing and managing information. The services and data include all the information services, servers, and data standards that collect, process, and store information. This includes the servers, data storage and distribution, cloud and edge servers, and software and data standards that allow the display of a [Common Operational Picture] COP as meaningful visual images that directly impart knowledge and increase understanding.”⁹ The fundamental problem with the implications of this section is that it seemingly eliminates the human element. The described equipment does not constitute a magical process that always works.

An important takeaway of this section is the implication that information services completely automate the data to information process; seemingly alleviating Soldiers from responsibilities for everything from data collection to information storage. However, based on training and the current state of Soldiers’ knowledge, Soldiers know that this is not reflective of reality. For example, if a Platoon Leader walks a motor pool with a green

notebook and writes down which vehicles have four seats, versus which vehicles only have two seats, then computes a total number of seats available, he or she has fundamentally completed the data to information process in a purely analog manner.

As referenced in ADP 6-0, a Clausewitzian assertion stands in stark contrast: “War is a human endeavor . . . not a mechanical process that can be precisely controlled by machines, calculations, or processes.”¹⁰ Humans can use technology to process data into information, but information services are neither a fix-all nor a completely autonomous process. Humans have to ask the right questions and know how to process data to create relevant information. This doctrinal review analysis concludes with the same original question – how does the Army process data into information?

Review Analysis of Tactical Doctrine: FM 6-0 Commander and Staff Organization and Operations

The above review analysis prompts the continued inquiry of accurately understanding the way the Army defines the term processed. FM 6-0 provides the Army’s definition. For context, FM 6-0 references data 111 distinct times, whereas the entire section that describes this process is only 55 words;¹¹ clearly indicating that the Army recognizes that data is important while the process of turning it into information is less-important. This idea relates to this research paper’s original food metaphor to ask, “is the act of cooking more or less relevant than the raw ingredients?”

“Processing information describes the act of cataloging data via established and usually routine sets of procedures to link or fuse it with other related data in order to create information. Information dissemination management and content staging capabilities enable the automated registering of data in order for it to be developed into

information and stored until needed.”¹² The answer to the question of “how does the Army process data into information,” becomes clear. The Army processes data “via established and usually routine sets of procedures” which narrowly “link or fuse [data] with other related data in order to create information.”¹³ This is a wholly unsatisfactory answer which again implies that this is an automated process. If this is the case, readers must assume that the process is further clarified within the context of information management. How does the Army envision this automated process happening?

Review Analysis of Signal Doctrine: FM 6-0.2
Signal Support to Operations

FM 6-0.2 initially articulates that “Signal personnel conduct information management” and lists eight specific tasks that they are responsible for, (1) application and database administration, (2) data backup and migration, (3) website interface maintenance, (4) troubleshooting, (5) security, (6) configuration, (7) providing network architecture, and (8) technological tools (to support content management and content sharing, and providing Department of Defense Information Network (DODIN) operations and information management support through the G-6/S-6 (IT) section.)¹⁴ These eight tasks are foundationally associated with keeping networks operational and secure. None of the eight tasks address how the Army collects data (as stated in the definition for Information Management) or, more importantly to this inquiry, how the Army processes data into information.

FM 6-0.2 further refines the responsibility for information management. “The network enterprise center (NEC) provides overall [Department of Defense Information Network] DODIN operations, and is responsible for information management,

information technology management, and telecommunication services.”¹⁵ This further indicates the assumption that operationally or tactically assigned Signal personnel are alleviated from information management responsibilities in garrison. For example, under this logic, if information management is the fundamental process that turns data into information, and the NEC is responsible for this process in garrison, a reasonable conclusion might be for Commanders to call the NEC to complain when Command and Staff slides are incorrect.

Doctrinally, information management is seemingly defined differently in FM 6-0.2 than in FM 6-0. In general, Signal personnel are assigned responsibility for information management, but the term is narrowed to no longer include the process of turning data into information. Soldiers should know that the process of turning data into information is a shared responsibility. Officers in all formations, at every echelon, routinely turn data into information.

Cognitive Processes: Cognitive Techniques in Visual Data Interpretation

For this analysis, one relevant scientific model is the Information-Matter-Energy-Intelligence model. This theory states that the natural world is subdivided into two categories. The *tangible* (or the “physical world”) consists of both matter (food) and energy (light), as well as the *intangible* (or the “abstract world,” such as math). The physical and abstract worlds exist independent of humans. Humans use intelligence to “combine information elements that come from the abstract world, such as: data, information, knowledge, [and] experience.”¹⁶ In other words, humans use observations of the physical world (data) to create information through intelligence. This is depicted in Figure 9.

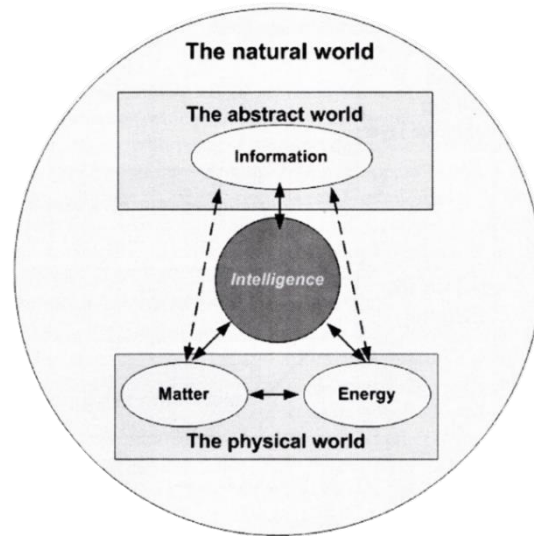


Figure 9. The IME-I Cognitive Information Model

Source: Lidia Ogiela and Marek R. Ogiela, *Cognitive Techniques in Visual Data Interpretation*, Studies in Computational Intelligence 228 (Berlin, Heidelberg: Springer, 2009), 53.

This leaves the discussion in largely the same place; wondering about the specifics of how data is turned into information. Exploring this concept further, three different specific information processing models stand out as relevant: cognitive mapping dependent on either (1) *categories*, (2) *features*, or (3) *processes*.¹⁷

Using experiential “intelligence,” it is simple to apply each of these three mapping techniques to typical Army information creation. A Platoon Leader (PL) walks the motor pool and uses categorical mapping to create a list of information: three High Mobility Multipurpose Wheeled Vehicles (HMMWV) and one Light Medium Tactical Vehicle (LMTV). The Company Executive Officer (XO) walks the same line and uses feature mapping to create a list of information: two up-armored, two-seater HMMWVs, one soft skin, four-seater HMMWV, and one soft skin, 3-seater LMTV with a cargo

capacity of an additional 12 Soldiers. A new Private uses process mapping to fill in a 5988 with information while conducting a weekly Preventive Maintenance Check and Service (PMCS): one headlight is broken.

Each of these mapping techniques (“intelligence”) is necessary in different contexts. Process mapping can standardize information creation to the lowest level, but may not always meet emerging requirements. The ability to adjust between each (critical thinking) is crucial when creating information for different decisions. Conversely, an overreliance on one mapping method (generally process mapping) can stovepipe thinking and limit an Officer’s ability to completely and accurately enable Commanders with the understanding that is required for their role in the operations process.

The above discussion has provided information regarding the Army’s expectation of data processing, the doctrinal definitions and how the Army defines and values aspects of data processing, and how people cognitively conceptualize data to draw conclusions or understand the meaning depicted within data. To better understand the holistic concept of data processing, it is prudent to analyze the civilian counterpart representation of knowledge and view of data processing. Similar issues with data and information must exist outside the Army, which begs the question of what educational “category” should this research address to make an accurate comparison? The following section will explore how civilian institutions view data processing and how they teach data.

Civilian Equivalence: How Many Ways Can We Teach Data Literacy?

After beginning this inquiry, focusing on the field of data visualization, a broader and more applicable category quickly emerged. *Data literacy* is a relatively new umbrella term that groups a number of closely related subjects. Because of the recency of data

literacy, much of the applicable civilian literature can be found by searching for the more precise term *information literacy*. As articulated here, information literacy as a category existed before the prevalence of high-performance computing. As the information environment changed, our vernacular had to follow, and data literacy evolved as a subset of information literacy.¹⁸

The source “How Many Ways Can We Teach Data Literacy?” goes on to offer a very broad conception of what data literacy is, which provides a useful starting point. “Data literacy is, first of all, critical thinking applied to evaluating data sources and formats, and interpreting and communicating findings.”¹⁹ When discussing the Army’s process of turning data into information, this seems to be what doctrine is suggesting. It ties the human element of critical thinking (*intelligence*) to sources/formats (analog vs digital, trusted vs non-trusted, etc.), interpreting (the application of information processing models), and communicating findings (providing a Commander an understanding of the OE to facilitate their role in the operations process).

In relation to data processing and information literacy, this study identified data literacy as the civilian educational category that encompasses the intelligence necessary to turn data into information, which includes the necessary mental processes as well as the technical knowledge necessary to complete the Army’s doctrinal requirement. Taking a holistic view of this process, individuals must be considered. This research further investigates the question of whether this intelligence is already present? Do Officers commission with the fundamental intelligence in data literacy required to process data into information? If so, this theoretically would alleviate the responsibility to either doctrinally articulate the process, or educate the force on best practices. To address these

questions, civilian research adds to the empirical data that strengthens the baseline of this discussion.

Current State (Civilian): Lead with Data

Qlik is an American software company that commissioned a research survey in 2018 which attempted to identify the current state of data literacy in the civilian world.²⁰

Their research produced four key statistical findings that are relevant:

First, *only “21% of 16-24-year-olds are data literate, suggesting schools and universities are failing to ensure students have the skills they need to enter the working world.”*²¹ [emphasis added] This finding suggests that growing up in a world saturated with software applications does not translate to a fundamental understanding of how software uses data to produce information. Deep familiarity with the user interface of Instagram, Facebook, or YouTube might allow someone to create a viral video, but these skills do not directly translate to someone’s ability to enable the Commanders role in the operations process.

Second, *only “24% of business decision makers surveyed are fully confident in their ability to read, work with, analyze, and argue with data.”*²² [emphasis added] This suggests that merely working in an environment that uses data does not broadly translate into proficiency with the resource. Similarly, working in a restaurant as a waiter or dishwasher does not somehow equate to cooking proficiency. Some individuals seek out relevant education and training when placed in environments that encourage this type of proficiency. These forward-leaning individuals may account for the three% difference between this statistic and the previous one.

Third, “85% of data literate people say they are performing very well at work, compared to 54% of the wider workforce.”²³ [emphasis added] This finding suggests that data literacy increases work performance, or at very least, self-confidence in the workplace. In today’s interconnected and data-saturated environment, it is simple to see how this talent translated into performance.

Finally, “78% of business decision makers said they would be willing to invest more time and energy into improving their data skillset.”²⁴ [emphasis added] This statistic suggests that there is a perceived need for these talents, despite not being widespread. It is easier to encourage individuals to learn if they are motivated to absorb the content.

These statistics clearly show that Officers do not commission with the fundamental knowledge of how to process data into information. Thus far, this research has shown that 1) Officers do not commission with these skills, and 2) the civilian educational equivalency that addresses the abilities that Army doctrine requires is entitled “data literacy.” The next logical step is to identify a solid definition for data literacy and further expand on concepts and their applicability within the Army.

Deep Dive: Be Data Literate

Due to the relative newness of the term “data literacy,” a number of different definitions currently coexist. This research has identified one definition that articulates the doctrinal requirement to process data into information and can provide the Army with a useful framework: “Data literacy is the ability to read, work with, analyze, and communicate with data.”²⁵ This description provides a simple and understandable process that can set the stage for future development.

To emphasize why this conception of data literacy is important, this literature offers a statistic that, if directly applicable to the military, should be concerning, “*Over a third of the study’s participants said they would find alternative methods to completing a task instead of using data, and 14 per cent said they would avoid it altogether.*”²⁶

[emphasis added] Commanders make life and death decisions based on their understanding of the OE. If one-third of their staff has avoided using an incredibly relevant and available resource due to a lack of ability to read, work with, analyze, or communicate with data, it will assuredly lead to otherwise avoidable gaps in a Commander’s understanding. Imagine the quality of a meal produced by a chef that “finds alternative methods” or altogether avoids using raw ingredients.

In addition to the definition of data literacy, the “four levels of analytics” is a framework that can help readers conceptualize the “intelligence” needed for Soldiers to competently perform their doctrinal task of turning data into information.²⁷ First, *descriptive* analytics refers to looking back at what has happened.²⁸ The Army generally operates at this level when looking at reports such as the Equipment Status Report (ESR), which purely analyzes the facts of the past. The second level of analytics is *diagnostic*, which adds the “why” to the description.²⁹ For example, if a Battalion XO briefs that the ESR reflects a large amount of broken equipment due to a recent exercise or deployment, he or she is using diagnostic analytics. The third level is *predictive* analytics, which takes our analysis into the future.³⁰ This is the first level of analytics that allows humans to be proactive instead of reactive, which is exceedingly important in an organization such as the Army. An example of this would be if our XO reviews the long-range training calendar and predicts that there will be an uptick in maintenance requirements next

quarter due to an exercise. Finally, the fourth level is *prescriptive* analytics, which directs current actions based on predictions.³¹ Imagine if the same XO identifies the top five long lead time parts that are most likely to fail in the upcoming exercise, and preemptively orders the equipment with enough lead time so that they are on-hand immediately following the exercise.

Each subsequent level of analytics requires increasing amounts of computational power and human input. Limiting this analysis to the first two levels of analytics, our previous example of descriptive analytics only required the Global Combat Support System – Army (GARMY) database to identify the current ESR. To move to diagnostic analytics, one must align the historic training calendar from the Digital Training Management System (DTMS) with the dates that parts were ordered in GARMY. Getting these non-interoperable systems to interface in a situationally dependent way is a technical task, one which either requires data literacy or must be done manually with varying levels of accuracy. For example, a quick manual first glance might indicate that the exercise caused the equipment failure, but if the majority of parts were ordered the week before the exercise instead of the week after, it could indicate a more systemic problem with the quality of routine PMCS; different problems, which require different solutions. Moving from diagnostic into predictive or prescriptive analytics only increases the number of variables and correspondingly increases the data literacy “intelligence” required to perform accurately.

Seemingly channeling Clausewitz, this literature goes on to repeatedly emphasize that “software and technology are here to help augment humans, if we are trained and educated enough to let it happen.”³² As war is a human endeavor, the Army must rely on

the Soldier's ability to understand and execute these levels of analytics rather than expect automated solutions to compensate for any potential wartime eventuality. Envisioning a world where each of our currently stove-piped systems are re-designed to fluently communicate, Murphy's law will continue to affect everything. Furthermore, our peer adversaries are afforded significant opportunities if the Army solely relies on automation to execute higher levels of analytics. AI and ML are wonderful innovations that may be able to assist in this regard, but they are not a panacea of situational understanding without an educated and data-literate Soldier working in concert with this technology.

Once convinced that information creation and analytics require a human-in-the-loop, the next questions take shape: does the Army do anything with this information? Is there a way of training our workforce to operate this way? Is data literacy education the correct approach? One way to answer these questions is to examine if civilian organizations have successfully implemented data literacy training.

Civilian Training: Teaching A One-Credit Course on Data Literacy and Data Visualization

As part of this research study, it is important to understand what training is already present. Most Officers graduate from civilian institutions prior to commissioning, so this training is relevant to their understanding of data. Unfortunately, in undergraduate institutions, "there is no one uniform approach to ensure consistency of instruction . . . data literacy is taught sporadically based on demand."³³

One advantage the Army has over civilian institutions is the existence of a robust PME system that can ensure a single uniform approach to data literacy education. As the civilian sector may struggle with education standardization, Army institutions alleviate

some of these issues and present a relatively simple method for implementing this type of training. That being said, this type of curriculum will only be added if it is worth the investment. The objective of this research methodology is to design a study that will determine what benefits may arise from this type of curriculum.

Summary

Beginning our Army-specific data literacy journey with doctrine, this research identified a few key take-a-ways. Data is an important concept in our doctrine, which clearly and repeatedly discusses that data is processed into information in order to provide Commanders with an accurate understanding of the operational environment.³⁴ Unfortunately, the term *processed* is vague and sometimes handwaves the entire process as automated. Attempting to identify what this process actually entails, the Information–Matter–Energy–Intelligence cognitive information model is used to describe how intelligence, specifically three different cognitive mapping processes, enables human brains to produce information.³⁵ Data literacy is then identified as the educational category that addresses our doctrinal gap, which is exasperated by the current lack of undergraduate data literacy training. Finally, this research outlined fundamental data literacy concepts and applied them to the Army.

¹ Headquarters, Department of the Army (HQDA), Army Doctrine Publication (ADP) 5-0, *The Operations Process* (Washington, DC: Army Publishing Directorate, 2019), 2-17.

² HQDA, ADP 6-0, 2-14.

³ *Ibid.*, 2-4.

-
- ⁴ Ibid.
- ⁵ Ibid.
- ⁶ Ibid.
- ⁷ Ibid., 3-8.
- ⁸ Ibid.
- ⁹ Ibid., 4-8.
- ¹⁰ Ibid., 1-1
- ¹¹ Headquarters, Department of the Army (HQDA), Field Manual (FM) 6-0, *Commander and Staff Organization and Operations* (Washington, DC: Army Publishing Directorate, 2014).
- ¹² Ibid., 3-9.
- ¹³ Ibid.
- ¹⁴ Headquarters, Department of the Army (HQDA), Field Manual (FM) 6-02 *Signal Support to Operations* (Washington, DC: Army Publishing Directorate, 2019), 2-23.
- ¹⁵ Ibid., 2-29.
- ¹⁶ Lidia Ogiela and Marek R. Ogiela, *Cognitive Techniques in Visual Data Interpretation*, Studies in Computational Intelligence 228 (Berlin, Heidelberg: Springer, 2009), 53.
- ¹⁷ Ibid., 36.
- ¹⁸ Yun Dai, “How Many Ways Can We Teach Data Literacy?” *IASSIST Quarterly* 43, no. 40 (January 2, 2020): 1–11.
- ¹⁹ Ibid., 1–11. 2.
- ²⁰ “Lead with Data: How to Drive Data Literacy in the Enterprise,” Qlik, 2018, https://www.qlik.com/us/bi/-/media/08F37D711A58406E83BA8418EB1D58C9.ashx?ga-link=datlitreport_resource-library, 4.
- ²¹ Ibid.
- ²² Ibid.
- ²³ Ibid.

²⁴ Ibid.

²⁵ Jordan Morrow, *Be Data Literate* (London: Kogan Page, Limited, 2021), 36.

²⁶ Ibid., 8.

²⁷ Ibid., 18.

²⁸ Ibid., 20.

²⁹ Ibid., 23.

³⁰ Ibid., 26.

³¹ Ibid., 29.

³² Ibid., 9.

³³ Tatiana Usova and Robert Laws, “Teaching a One-Credit Course on Data Literacy and Data Visualisation,” *Journal of Information Literacy* 15, no. 1 (January 2021): 84–95. <https://www.ariadne.ac.uk/JIL/article/view/PRJ-V15-I1-3>.

³⁴ HQDA, FM 6-0.

³⁵ Ogiela and Ogiela, *Cognitive Techniques in Visual Data Interpretation*, 36.

CHAPTER 3

RESEARCH METHODOLOGY

Introduction

Applying the concepts articulated in the preceding literature review, this methodology is designed to provide a diagnostic analysis of the current state of data literacy in the Army. As the hypothesis reasonably correlates as expected, statistical models elevate the research to predictive analysis. Properly communicated, these results can help Commanders visualize the current force's understanding of data literacy.

Restated Research Questions

Two categories were considered: the input and the output. The input consists of two elements: characteristics required of officers (drawn from Department of the Army Pamphlet 600-3 Smartbooks), and institutional training (drawn from a review of the Basic Officer Leader Course Programs of Instruction). The output consists of a survey targeting the Command and General Staff Officer College (CGSOC) population, which addresses elements such as respondents' perception of their individual understanding of data. As the CGSOC student population holds the rank of either senior Captain or junior Major, this survey will specifically address the "mid-grade" Officer population. This analysis will be conducted to answer the primary research question of:

Primary Research Question

Is there a significant difference in the average amount of time mid-grade U.S. Army Officer's spend manipulating data or analyzing information comparing across BOLC attended or current branch?

Four secondary research questions will be explored in order to provide clarity and depth to this analysis:

Secondary Research Questions

1. How much does Professional Military Education data literacy training change the amount of time mid-grade U.S. Army Officers spend manipulating data?
2. How much does Professional Military Education data literacy training change the amount of time mid-grade U.S. Army Officers spend analyzing information?
3. How much does Professional Military Education data literacy training change a mid-grade U.S. Army Officer's familiarity with Excel?
4. How much does Professional Military Education data literacy training change a mid-grade U.S. Army Officer's perception of their own data literacy?

Method

This study uses multi-methodology techniques. It consists of a data review, quantitative in nature, composed of three data sources: (1) a survey of the CGSOC class of 2022, (2) Department of the Army Pamphlet (DA PAM) 600-3 branch Smartbooks, and (3) Basic Officer Leadership Course (BOLC) Program of Instructions (POIs). On their own, each dataset constitutes a descriptive analysis of the current state of Army data literacy education. The combination of the data elevates the analysis to the diagnostic level. The rationale for the type of methodology is to triangulate the literature review, theory, and data from the field.

The survey is designed to descriptively analyze the current state of data literacy in the Army. Likert scale results will allow researchers to quantify and assess respondents'

familiarity and comfort with data literacy skills. The DA PAM 600-3 review is a descriptive analysis of pre-commissioning requirements. By scanning for the word “data,” this analysis will quantify the number of times each branch references this fundamental skill within the “required characteristics” section of these documents in order to measure the emphasis each branch places on data when recruiting candidates.

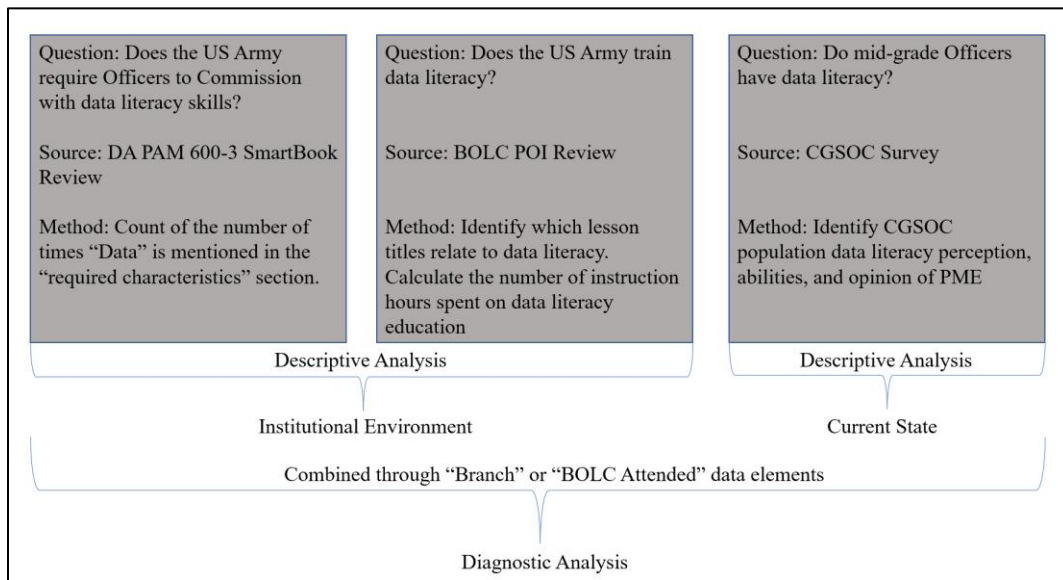


Figure 10. Methodology Logic Chart

Source: Created by author.

The BOLC POI document review is a descriptive analysis of the Professional Military Education (PME)/institutional training environment. By aligning lesson titles with their associated Instructor Contact Hours (ICH) within each specific branch, and identifying data literacy relevant classes, this research will generate a cumulative assessment of relevant course hours. Once complete, it will be possible to align the POI results by “branch” with the survey “BOLC attended” data to determine correlation

between data literacy training and survey responses. This alignment elevates each individual diagnostic analysis into a descriptive analysis intended to articulate how the Army got to the current state of data literacy (Figure 10).

To achieve this level of analysis, Microsoft Power BI is used to create a simple data model which combines each dataset. Lookup tables are used to align columns with data tables through the use of primary and foreign keys and relationships (one to one and one to many). The data model is depicted in Figure 11, which identifies the relevant keys and corresponding relationships. Lookup tables are designed to 1) reduce the size of data tables, and 2) allow data visualization software to relate data tables to answer specific queries. Digital copies of these lookup tables are available upon request. The “measures” feature is used to explore the data and calculate insights before addressing the research questions through statistical tests.

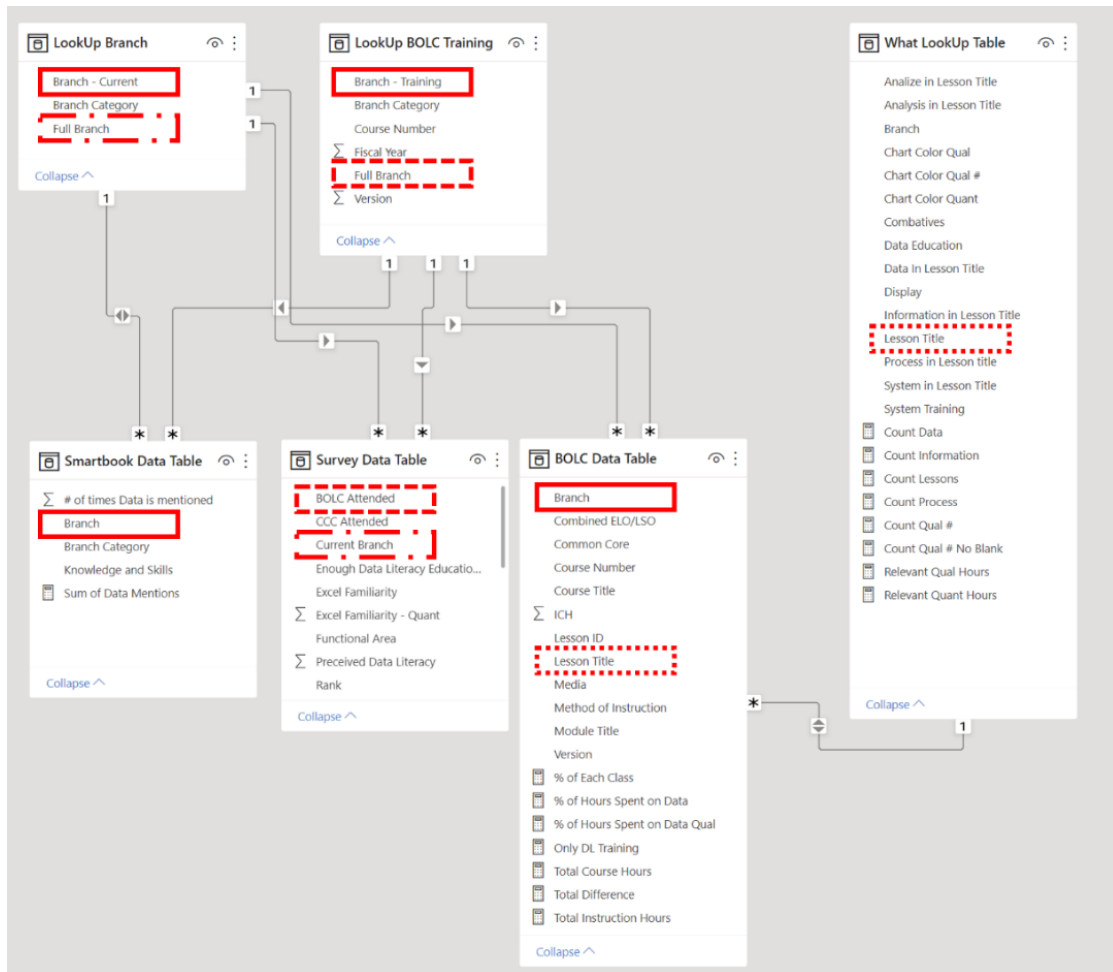


Figure 11. Data Model

Source: Created by author.

Multi-methodology was selected as opposed to mixed-methodology by virtue of data groups involved in the research. For example, there is a distinct difference between mixed methodology and multi-methodology. The virtue of mixed methodology includes collecting both qualitative and quantitative data. The virtue of multi-methodology includes the collection of multiple types of qualitative or quantitative data but not a mix of qualitative and quantitative (can be qualitative/qualitative, or quantitative/

quantitative).¹ This research only collected quantitative information, but qualitative techniques were used to determine relevancy.

To accurately and thoroughly analyze and interpret the data, each resource category initially used exploratory analysis. Each method will follow the same macro-order: identify the relevant data source, access the data source, clean the raw data, analyze the clean data, and visualize the resulting information. In order to produce a repeatable process for future research, each specific method will be listed below.

Data Collection

CGSOC Class of 2022 Survey

The rationale for conducting a survey is to gather opinions of a distinct population that possesses experience with Army Data Literacy training and job-specific experience with data processing, as used in the Army. This survey population is able to examine a gap in understanding through the consolidation of a dataset that examines Army Officer's current sentiment and educational experience within the field of data literacy. As this data did not exist prior to this survey, to the knowledge of the researcher, it required original research. The final survey questions can be reviewed in Appendix A. Upon completion of the survey, all documents were scanned for personal information, which was immediately destroyed by the Quality Assurance Office (QAO). The data was then cleaned and transformed into a relevant data set, as can be seen in the data model in Figure 11.

The explicit steps taken during survey data collection included: (1) the researcher determined the population as convenience sampling, (2) the type of survey as Likert or nominal questioning using the TRADOC approved survey system VERINT, (3) survey questions which were designed to specifically answer the research questions, (4) received

CGSOC Institutional Review Board (IRB) approval, (5) distributed the survey using the approved channels within the organization, (6) ensured the safeguarding of all anonymized survey materials, (7) acknowledged and abided by the Belmont Report code of ethics, (8) analyzed the data, and (9) assembled a findings report write-up.

Determining the Population and Sample

Convenience sampling was selected for this research. This sampling technique was chosen due to the geographic proximity, ease of sampling, and the experience base of the CGSOC student population. Furthermore, the student population fits the inclusion criteria of mid-grade, United States Army, Officer.² The CGSOC class of 2022 consists of 990 total U.S. students, 780 from the United States Army (active duty). Power analysis indicates that with a confidence level of 95% and a margin of error set at 5%, 258 survey participants are required. In order to facilitate this response rate, the researcher invited 774 students to participate in the survey.

Determine the Type of Survey

The type of survey selected for this part of the research is a quantitative scale measure survey consisting of 10 questions and included a front-page with a control number, welcome, and necessary disclosure statements. Participants had the ability to opt out as the survey was not mandatory. Questions used the Likert scale or nominal scale. The system that deployed the survey was the TRADOC approved survey system VERINT.

Designing the Survey Questions

The design of the survey specifically answered the research questions and were purposely constructed as to not include any personal information. The questions were educational and opinion in nature and did not pose mental or emotional prompts that would occur above what is considered a normal baseline (questions listed in Appendix A). Survey questions went through several iterations and were vetted by three Ph.D. volunteers within TRADOC, having knowledge of research methodology and data literacy.

Receive CGSOC Institutional Review Board (IRB) Approval

CGSOC has a Human Protections Director responsible for ensuring the proper review and conduct of research involving human subjects. The proper procedures, as guided by the office, were followed precisely and all training required (Collaborative Institutional Training Initiative, CITI) was attained prior to any research with human subjects.

Distribute the Survey

The distribution of the survey was under the control of the CGSOC Quality Assurance Office (QAO). The survey opened for an agreed upon time, and closed after receiving the required number of participants. QAO de-identified all data before delivering it to the researcher.

Safeguard all Survey Materials

While the survey did not generate any personal information from the participants, it is still prudent to keep all aspects of the survey within the confines of the research and

to those who have a need to know. The survey is on available to those who are identified as part of the research process. The electronic data will be secured on a Common Access Card (CAC) accessible computer. The computer was safeguarded by personal protection while working on the research, and by locking the computer in a secure room/desk when the research was not being worked on.

Analyze the Clean Data

Independent variables allow the dependent variables to be analyzed for relevant connections. To this end, the research collected the following independent variables: rank, current branch, previous branches, BOLC attended, CCC attended, and Reserve CCC (yes/no). Dependent variables are the time spent manipulating data, the time spent analyzing information, sufficient PME training, PME perception, and Excel familiarity. Once complete, the survey data was be inserted into the data model (Figure 11), which enabled exploratory analysis and relevant statistical tests.

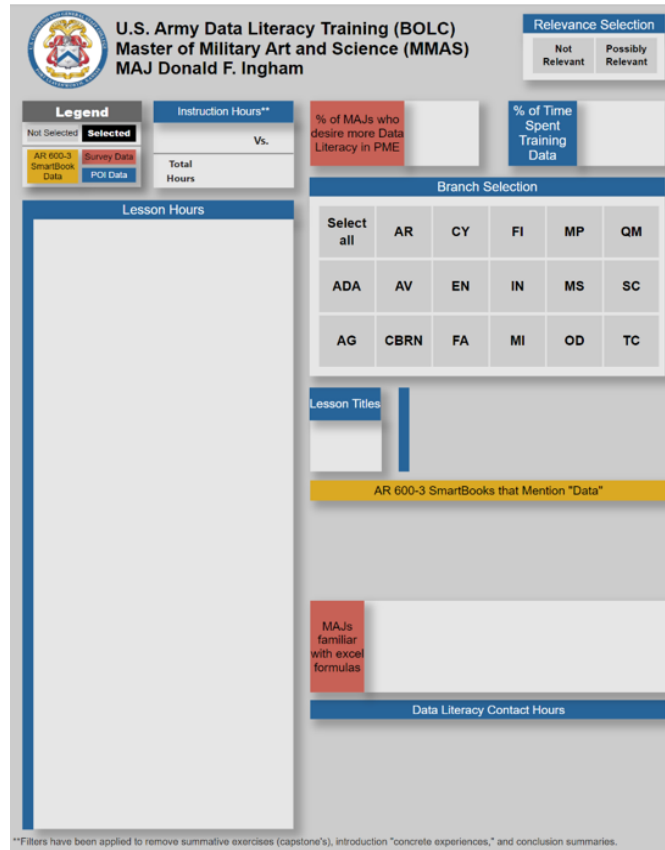


Figure 12. Base Visualization

Source: Created by author.

Visualizing the resulting information – upon completion of the analysis phase, “cards,” “charts,” and “tables” were used within the base Power BI visualization (Figure 12). The visualization dynamically filters relevant data to help communicate relevant information and insights. Survey results are depicted in red. A copy of the relevant .pbix file is available to relevant researchers upon request (CUI classification level).

DA PAM 600-3 Smartbooks

Identifying the relevant data source – in order to answer the question of whether each branch requires Officers to commission with data literacy skills, the researcher

chose to analyze the “Officer characteristics required” section of each DA PAM 600-3 branch Smartbook. Access the data –this information was readily accessible through MilSuite at <https://www.milsuite.mil/book/groups/smartbook-DA-PAM-600-3>. Clean the raw data – after consolidating the relevant documents, the data was transformed to fit into the data model (Figure 11).

Analyze the Clean Data

Data was grouped as either independent or dependent. Independent variables include: branch, branch acronym, and branch category. The only dependent variable is the number of times “data” is mentioned within the “required characteristics” section of each Smartbook. A chart was created with the X-axis being the branch and the Y-axis visualizing the number of times the term “data” is present. No further analysis of this dataset was necessary, as these variables are not needed to prove the research questions. These figures are calculated to serve as an indicator, not as proof of a correlation. Visualizing the resulting information - upon completion of the data analysis phase, this chart was inserted into the base Power BI visualization. The visualization dynamically filters relevant data to help communicate relevant information and insights. Smartbook results will be depicted in gold.

Basic Officer Leadership Course (BOLC) Programs of Instruction (POIs)

Identify the relevant data source – to create an accurate picture of the lessons Professional Military Education (PME) provides to BOLC students, this research consolidated a table of every lesson taught and its corresponding Instructor Contact Hours (ICH). This data exists on each individual branch Program of Instruction (POI),

which are stored within a system called the Training Development Capability (TDC). Accessing the data – after receiving permission to use this data from the director of this course (Appendix E), Dr. Lynn Soots, research committee member, assisted in the consolidation of these documents. The “POI Summary Report” option consolidated the needed information. A listing of each POI analyzed can be viewed in Appendix D.

Cleaning the raw data – unfortunately, the only data format the researcher could find was .PDF (naming convention: Branch_Acronym Summary.pdf). Cleaning this data took a substantial amount of time and consisted of the following steps. The researcher initially converted each PDF file into Excel (naming convention: Branch_Acronym Summary.xlsx) using the “Save to Other > Spreadsheet” feature in Adobe. Complicating this process, PDF sections that fall on page breaks often cause a conversion error (Figure 13).

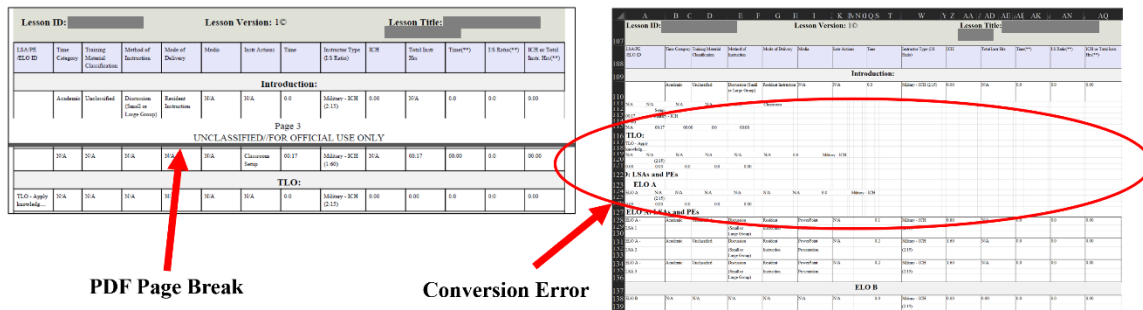


Figure 13. Conversion Issues

Source: Created by author.

In order to correct this issue, each Excel document was manually edited to ensure consistency (naming convention: Branch_Acronym Summary Fixed.xlsx). Once manual edits were complete, the researcher created a data normalization process in the Power BI

– Power Query Editor (File Name: 0. Data Cleaning.pbix), with four standardized steps (Figure 14).

To recreate this process: 1) Duplicate “template” query, 2) Change data source to the next branch “fixed” Excel document and change sheet name 3) Change the *first identified steps* code to the column number that actually displays the Lesson Title in the current dataset 4) Change *the second identified step* to the column that displays ICH 5) Change *the third identified step* to the column that displays MOI 6) Change *the fourth identified step* to the column that displays Media 7) Delete all columns left of “course #” 8) Rename and reorder the columns 9) Add a conditional column that determines if a module is Common Core or Branch Technical. Upon completion, export a consolidated database into its own Excel document (Combined BOLC Dataset.xlsx) in order to eliminate rendering time.

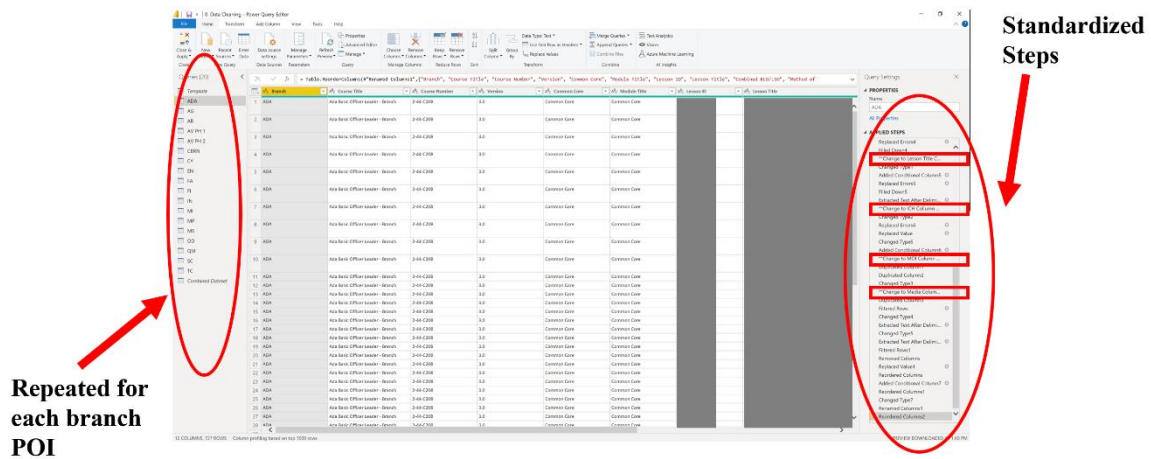


Figure 14. Power Query Editor Data Normalization

Source: Created by author.

Analyze the Clean Data

Independent variables are: branch, course title, course number, version, common core, module title, lesson ID, lesson title, and combined ELO/LSO. Dependent variables are the method of instruction, media, and ICH. The clean spreadsheet was added to the data model (Figure 11) which enables exploratory analyses and relevant statistical tests. Visualize the resulting information – upon completion of the data analysis phase, “cards,” “charts,” and “tables” were used withing the base Power BI visualization. The visualization dynamically filters relevant data to help communicate information and insights. POI data is depicted in blue.

Data Analysis Process

To begin exploratory analysis, this research calculated the total number of Instructor Contact Hours (ICH), by lesson title and terminal learning objective. To create a more accurate picture, non-instructor contact hours (for example, lesson introductions, summaries, exercise, and capstones) were filtered from the calculation to identify “instruction hours” rather than “total hours” (total hours will directly relate to the number of hours listed on each POI). Each lesson title is analyzed in two ways: quantitatively and qualitatively. Quantitative analysis consisted of scanning lesson titles for the words “data,” “process,” and “information.” If the title lists any of these keywords, the instructor hours are identified as “data literacy relevant.” Qualitative analysis instead calculated instructor hours by individual determination. The researcher looked at every lesson title and individually determined if the lesson seems to be relevant. This analysis produced a “possibly relevant” column, which serves as a means of filtering non-relevant contact hours. Each lesson is represented as an axis on a chart, with data-relevant titles

highlighted for readability. The relevant instruction hours is used to calculate a total% of time each branch spends addressing data literacy concepts. The charts and numbers dynamically filter based upon user-selected buttons representing each basic branch. The qualitative analysis base visualization can be viewed in Figure 12. A list of the measure names and their location in the visual can be found in Figure 15. A graph of measure names, corresponding to the table it is referencing and the code it is executing, is located in Appendix G.

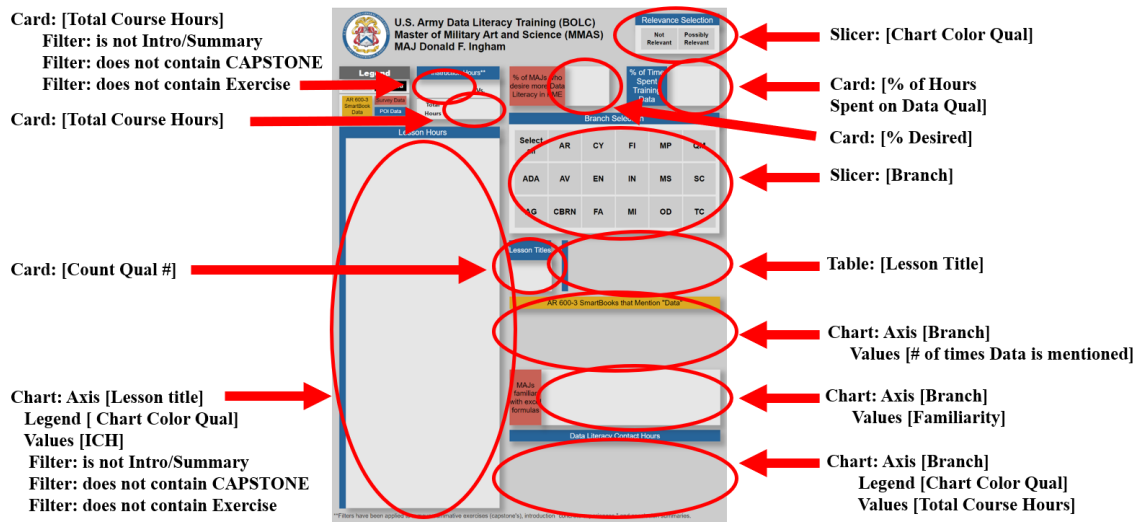


Figure 15. Data Analysis Measures

Source: Created by author.

Statistical Analysis

Upon completion of the data cleaning and exploratory analysis, the researcher conducted two different statistical tests. The first tests were two one-way MANOVAs, which addressed the primary research question. The second tests were four bivariate correlations, and potential corresponding bivariate regressions, which addressed the secondary research questions.

1. Primary Research Question (Figure 16 – Branch acronyms are listed in Appendix H):
 1. Question: Is there a significant difference in the average amount of time mid-grade U.S. Army Officer's spend manipulating data or analyzing information comparing across "BOLC attended" or "current branch?"
 2. Test: One-Way MANOVA #1
 3. Independent Variable: BOLC attended (from the survey)
 4. Dependent Variable #1: Time spent manipulating data (from the survey)
 5. Dependent Variable #2: Time spent analyzing information (From the survey)
 6. Looking for: P Value of ≤ 0.05 which would indicate a significant difference in average time between BOLC attended categories.
 7. Test: One-Way MANOVA #2
 8. Independent Variable: Current branch (from the survey)
 9. Dependent Variable #1: Time spent manipulating data (from the survey)
 10. Dependent Variable #2: Time spent analyzing information (From the survey)
 11. Looking for: P-value of ≤ 0.05 , which would indicate a significant difference in average time between current branch categories.

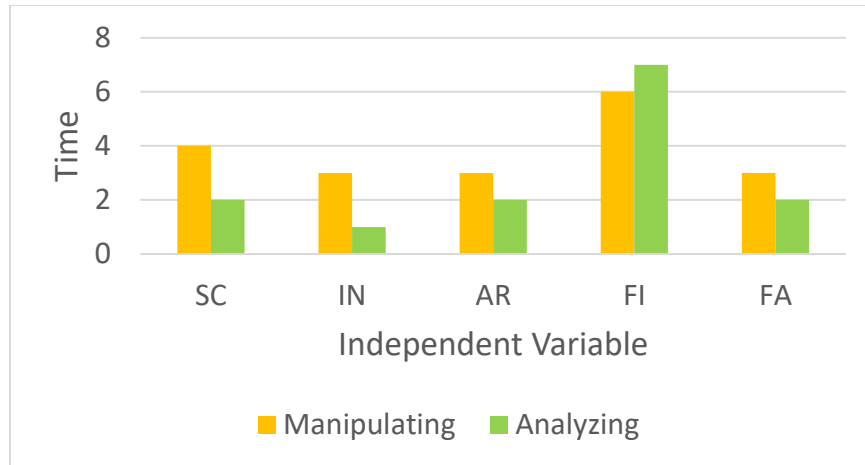


Figure 16. Example Primary Research Question Data Visualization (not actual data)

Source: Created by author.

2. Secondary Research Questions (Figure 17 – Branch acronyms are listed in Appendix H):

1. Question 1: How much does Professional Military Education Data Literacy training change the amount of time mid-grade U.S. Army Officers spend manipulating data?
2. Independent Variable: Instructor contact hours (from the BOLC POI, aligned through the “BOLC attended” question of the survey)
3. Dependent Variable: Time spent manipulating data (from the survey)
4. Looking for: Strong negative (R^2 of ≥ 0.5) correlation. If present, a regression line will indicate the required change in data literacy training required to achieve a corresponding change in the number of hours spent manipulating data.

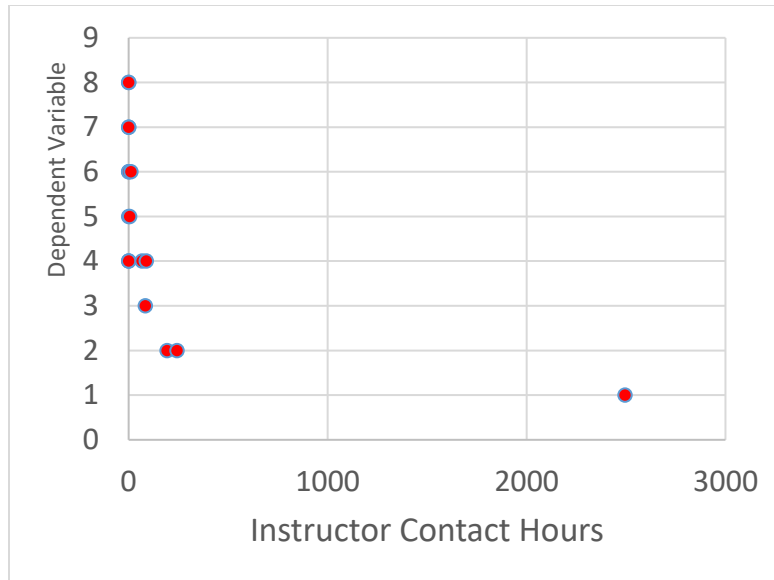


Figure 17. Example Secondary Research Question Data Visualization (not actual data)

Source: Created by author.

5. Question 2: How much does Professional Military Education Data Literacy training change the amount of time mid-grade U.S. Army Officers spend analyzing information?
6. Independent Variable: Instructor contact hours (from the BOLC POI, aligned through the “BOLC attended” question of the survey)
7. Dependent Variable: Time spent analyzing information (from the survey)
8. Looking for: Strong positive (R^2 of ≥ 0.5) correlation. If present, a regression line will indicate the required change in data literacy training required to achieve a corresponding change in the number of hours spent manipulating data.
9. Question 3: How much does Professional Military Education Data Literacy training change a mid-grade U.S. Army Officers’ familiarity with Excel?

10. Independent Variable: Instructor contact hours (from the BOLC POI, aligned through the “BOLC attended” question of the survey)
11. Dependent Variable: Excel familiarity (from the survey)
12. Looking for: Strong positive (R^2 of ≥ 0.5) correlation. If present, a regression line will indicate the required change in data literacy training required to achieve a corresponding change in Excel familiarity.
13. Question 4: How much does Professional Military Education Data Literacy training change a mid-grade U.S. Army Officers’ perception of their own Data Literacy?
14. Independent Variable: Instructor contact hours (from the BOLC POI, aligned through the “BOLC attended” question of the survey)
15. Dependent Variable: Perception of own data literacy (from the survey)
16. Looking for: Strong positive (R^2 of ≥ 0.5) correlation. If present, a regression line will indicate the required change in data literacy training required to achieve a corresponding change in perception.

Ethical Considerations

The underlying ethical foundation of this research follows *The Belmont Report*. The Department of Health, Education, and Welfare (1979) produced *The Belmont Report* as explicit guides for conducting research. This report focused on three key areas, respect for persons, beneficence, and justice.³

(1) Respect for persons: protecting the autonomy of all people and treating them with courtesy and respect and allowing for informed consent. Researchers must be truthful and conduct no deception. (2) Beneficence: the philosophy of “do no harm”

while maximizing benefits for the research project and minimizing risks to the research subjects. (3) Justice: ensuring reasonable, non-exploitative, and well-considered procedures are administered fairly – the fair distribution of costs and benefits to potential research participants – and equally.⁴ As part of the research, the chair of the research committee and the two committee members were consulted to ensure all elements of the research were appropriately conducted and ethical principles were followed.

This paper includes human-subject research in the form of a survey. In order to ensure this survey is ethical, the researcher will: (1) ensured that none of the questions could identify an individual, (2) all identifying information was immediately destroyed, and that (3) none of the questions placed respondents in an ethical quandary.

Furthermore, the researcher ensured that the email soliciting survey respondents clearly articulated informed consent, the intent of the survey, and possible repercussions of participation. Finally, the researcher will maintain the data on a jump drive within a personal document safe for a minimum of three years before the data is destroyed.

In order to address bias, the researcher heavily relied on the committee to formulate the survey questions. Numerous revisions were made including adding definitions, broadening questions, and ensuring that the participants have the ability to change answers prior to final submission. Similarly, the research questions and hypotheses underwent redundant committee reviews to reach consensus and eliminate bias.

Specific Ethical Assurances:

1. How were participants identified and excluded?

- a. Participants were identified through the VERINT system as current U.S. Army CGSOC students. Additionally, there were two survey questions that will allow researchers to ensure consistency. The first was the “current rank” question, which allowed the exclusion of any participants not within the rank of Captain or Major. Secondly, the “reserve component” will may allow an expansion of this research to encompass the Captain’s Career Course education system, while excluding participants that did not attend a resident course.
2. How were participants be recruited?
 - a. Participants were recruited through an email (Appendix B) sent by the CGSOC QA office, which requested students participate in the survey.
3. How many participants were involved? What is the rationale for that number?
 - a. The survey received 258 survey responses. This number was determined by the desire to have adequate representation across branches, ranks, and between the active versus reserve/national guard components.
4. How were conflict of interest mitigated?
 - a. Given the subject matter being researched, the likelihood of conflicts of interest were low. Despite this, the survey was completely anonymous, and all data points were given equal consideration in order to avoid skewing the resulting research.
5. How were participants informed of their rights and responsibilities?
 - a. Participants were informed of their rights within the email eliciting their participation in the survey.

6. What data was collected?
 - a. The data that was collected can be viewed in Appendix A, and the associated recruitment email and consent form can be viewed in Appendix B and C.
7. How was data be collected?
 - a. The CGSOC Quality Assurance (QA) office used the Army approved online survey system (VERINT) to conduct the survey. Once complete, the QA office de-identified the data before delivering it to the researcher.
8. How was data analyzed?
 - a. Data was analyzed using the techniques found with the preceding Data Analysis Process section.
9. How is data stored and secured (minimum of three years)?
 - a. Data is be stored on an external flash drive, which is secured in a fire/waterproof safe at the researcher's primary residence. It will remain here for no less than three years after the publication of this paper.

Summary

A multi-method data review was be used, relying on three different sources. (1) A survey amongst the CGSOC class of 2022, (2) A DA PAM 600-3 Smartbook document analysis and (3) A POI document analysis. These variables were compared using Power BI measures and appropriate statistical analyses. The results are visualized to improve information communication.

¹ John Creswell and Vicki Plano, *Designing and Conducting Mixed Methods Research* (Thousand Oaks: SAGE Publications, 2017).

² “Convenience Sampling: Definition, Advantages and Examples,” *QuestionPro* (blog), accessed March 29, 2022, <https://www.questionpro.com/blog/convenience-sampling/>.

³ Office of the Secretary, Department of Health, Education, and Welfare (HSS), *The Belmont Report: Ethical principles and Guidelines for the Protection of Human Subjects of Research* (Washington, DC: U.S. Department of Health and Human Services, April 18, 1979), <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html>.

⁴ Ibid.

CHAPTER 4

DATA FINDINGS AND ANALYSIS

Introduction

Applying the established methodology, this chapter interprets the collected data by processing it into visualized information. The intent of this research was to *describe* the current data literacy operating environment, *diagnose* the training that led to the current state, and (for statistically successful questions) *predict* changes in understanding based upon changes in training.

Restated Research Questions

Primary Research Question

Is there a significant difference in the average amount of time mid-grade U.S. Army Officer's spend manipulating data or analyzing information comparing across BOLC attended or current branch?

Secondary Research Questions

1. How much does Professional Military Education data literacy training change the amount of time mid-grade U.S. Army Officers spend manipulating data?
2. How much does Professional Military Education data literacy training change the amount of time mid-grade U.S. Army Officers spend analyzing information?
3. How much does Professional Military Education data literacy training change a mid-grade U.S. Army Officer's familiarity with Excel?
4. How much does Professional Military Education data literacy training change a mid-grade U.S. Army Officer's perception of their own data literacy?

Data Analysis

Exploratory Analysis Findings

CGSOC Class of 2022 Survey

The survey had a total of 258 respondents, meeting the Power Analysis participation number to indicate a 95% confidence threshold and a 5% error rate. Upon initial investigation, two respondents did not meet the “mid-grade Officer” requirement. The data associated with these participants was eliminated, resulting in the majority of the following analysis being conducted on the resulting 256 participants.

The first survey finding suggests that there is a high demand for data literacy training to be taught in PME: 89.68% of surveyed officers selected “No” when asked “Do you believe Professional Military Education (PME) has provided enough Data Literacy education/training (Data Literacy is defined as an ability to “read, work with, analyze and communicate with data”)?” These results showed no significant difference between BOLC attended (Figure 18). Branch acronyms are listed in Appendix H.

Is There Enough DL Training in PME?

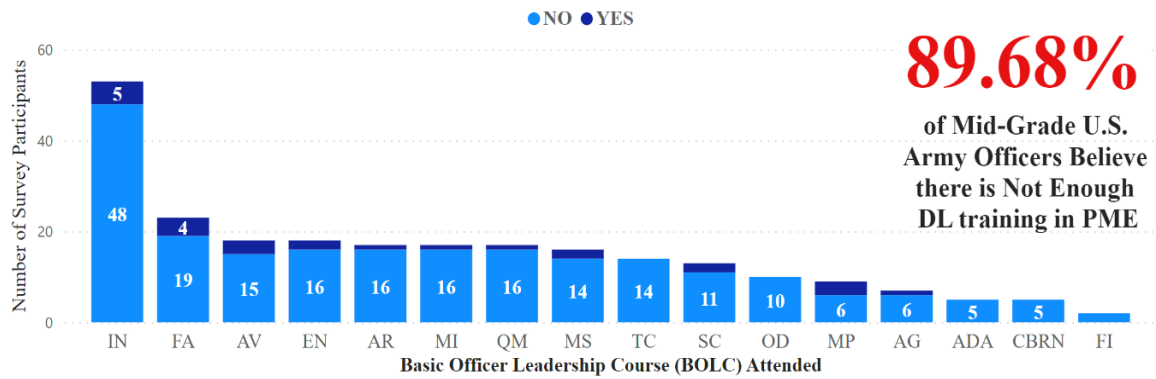


Figure 18. Mid-grade U.S. Army Perspective on Data Literacy Training in PME

Source: Created by author.

The second finding is that the average mid-grade US Army Officer spends an incredible amount of their time either manipulating data or analyzing information. This population is spending an average of 18 more minutes manipulating data than analyzing information every day (Figure 19). Assuming 260 annual work days, this calculates to 78 hours, or approximately two weeks every year, lost to manipulating data. For comparison, a National Guard Annual Training is two weeks.

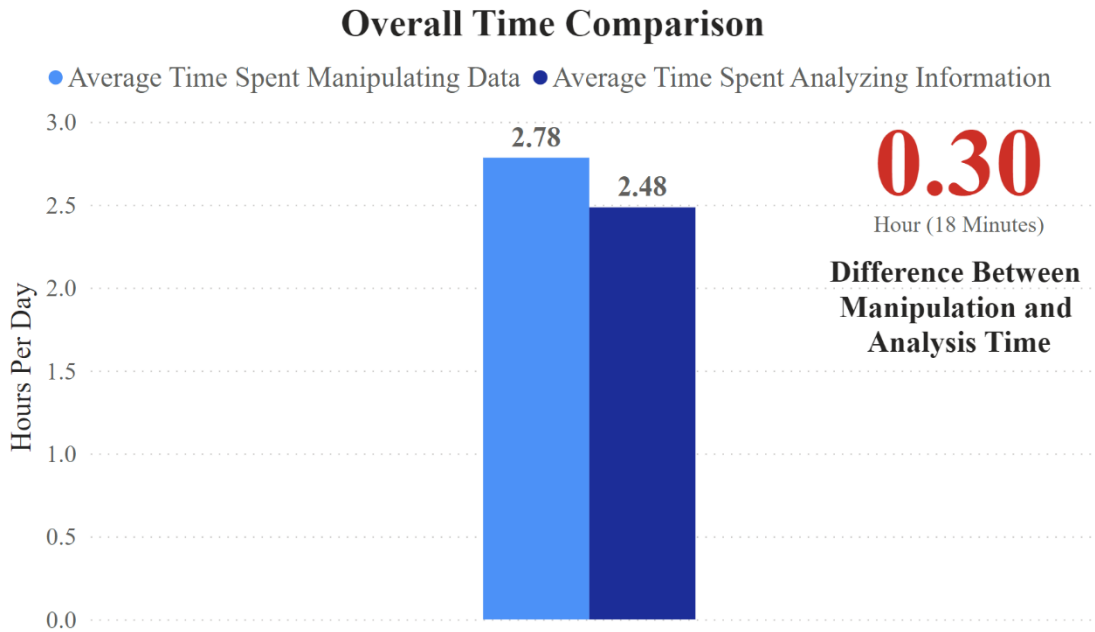


Figure 19. Average Total Time Comparison

Source: Created by author.

The third finding is that only Officers who have graduated from two BOLC courses (Military Intelligence and Chemical) spend more time analyzing information than manipulating data. This information is visualized in Figure 20. As sample sizes are important, this image includes the survey population size as an orange line. Branch acronyms are listed in Appendix H.

Figure 20. BOLC Impact on Time – Overall

Source: Created by author.

The fourth finding is that respondents' perception of their familiarity with Excel formulas does not correlate with institutional training (Figure 21). In this image, the gray

horizontal line indicates the number 3, which is “familiarity with Excel.” The columns below the line are considered “unfamiliar,” whereas the columns above are considered “very familiar.” Sample sizes are again depicted as an orange line. Branch acronyms are listed in Appendix H.

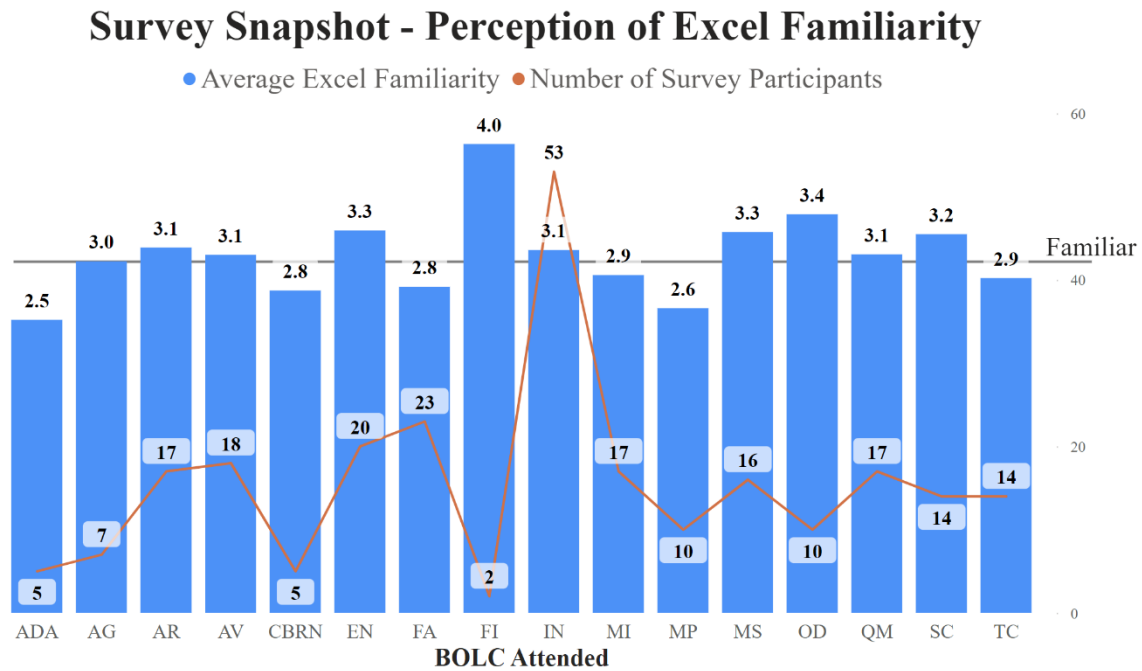


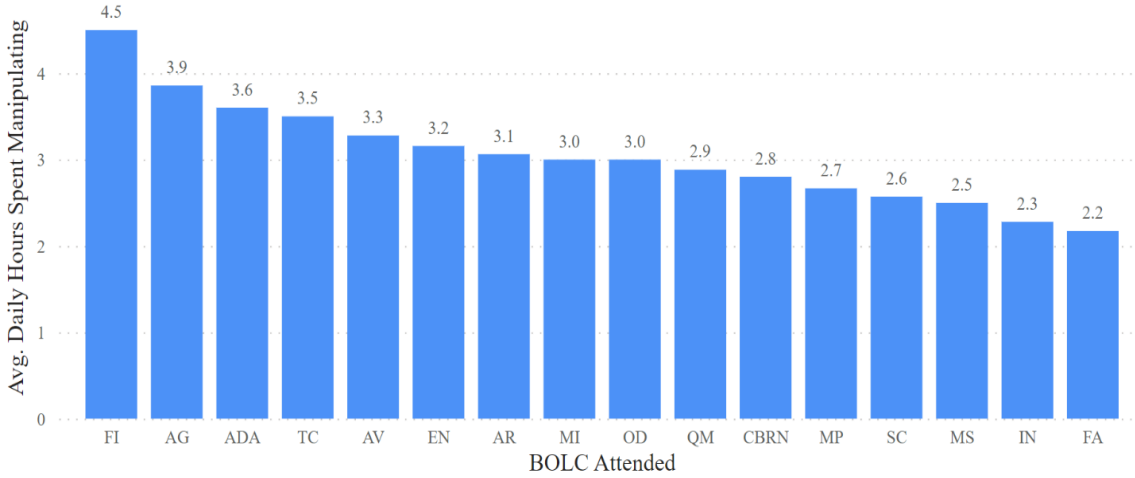
Figure 21. BOLC Impact on Perception of Excel Familiarity

Source: Created by author.

The fifth finding identified which BOLC graduates spent the most time manipulating data versus which BOLCs enabled their graduates to spend more time analyzing information. Finance graduates spend the most time manipulating data (4.5 hours per day), while Field Artillery graduates spend the least amount of time (2.2 hours per day) at this task. Conversely, Chemical graduates spend the most time manipulating data (3.6 hours per day), while Ordinance graduates spend the least amount of time

analyzing information (1.9 hours per day). This is visualized in Figure 22. Branch acronyms are listed in Appendix H.

Who Spends The Most Time Manipulating Data?



Who Spends The Most Time Analyzing Information?

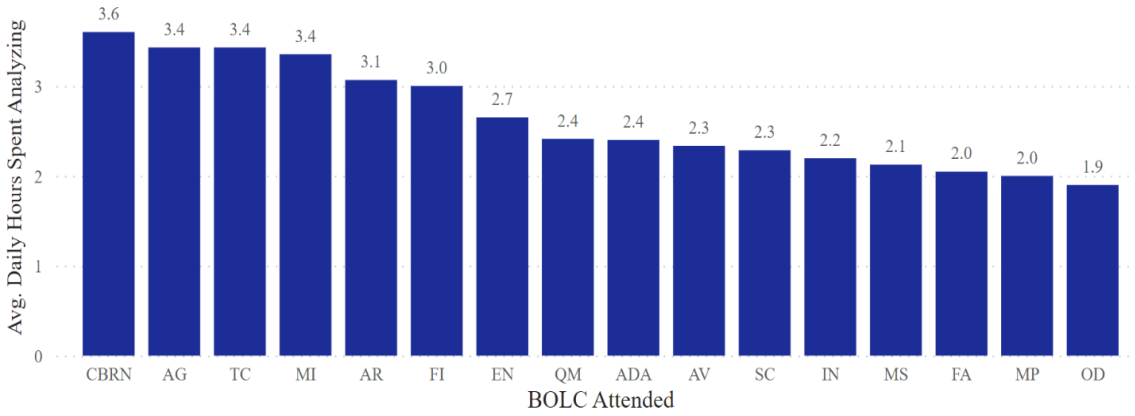


Figure 22. BOLC Impact on Time – Comparison

Source: Created by author.

DA PAM 600-3

The least useful analysis came from reviewing the “required characteristics” section of the Department of the Army Pamphlet 600-3 branch Smartbooks. Only two

branches (Finance and Aviation) mentioned “data” in this section. The finding from this review is that most branches do not require Officers to enter into the service having much proficiency, or interest, in data. This is visualized in Figure 23. Branch acronyms are listed in Appendix H.

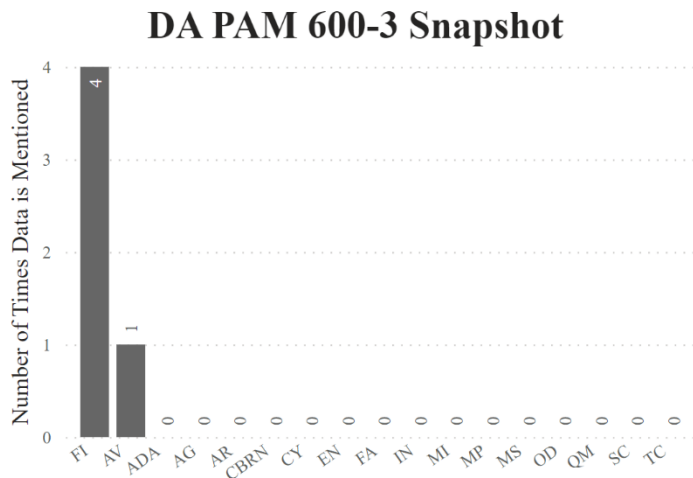


Figure 23. DA PAM 600-3 Data Importance

Source: Created by author.

Basic Officer Leadership Course (BOLC) Programs of Instruction (POI)

Exploratory analysis of the BOLC POIs began with the Quantitative approach to identifying data literacy relevant coursework. By scanning lesson titles for the “data,” “process,” and “information” keywords, this research found that seven lessons contained the word data, 29 lessons contained the word “process,” and an additional 15 lessons contained the word “information.” To maintain accuracy, this research applied a filter to remove non-instruction hours during all coursework (for example introductions,

summaries, capstone exercises, etc.). After filtering, this researcher is able to calculate the percentage of Instructor Contact Hours (ICH) associated with these classes. This analysis determined that the average Army Officer is trained on data literacy relevant coursework during 1.89% of their time at BOLC (Figure 24). Upon further analysis, the quantitative approach findings were discounted because a significant number of the lesson titles do not relate to data literacy. For example, many of the “process” lesson titles were in fact “in-processing” or “out-processing.” Branch acronyms are listed in Appendix H.

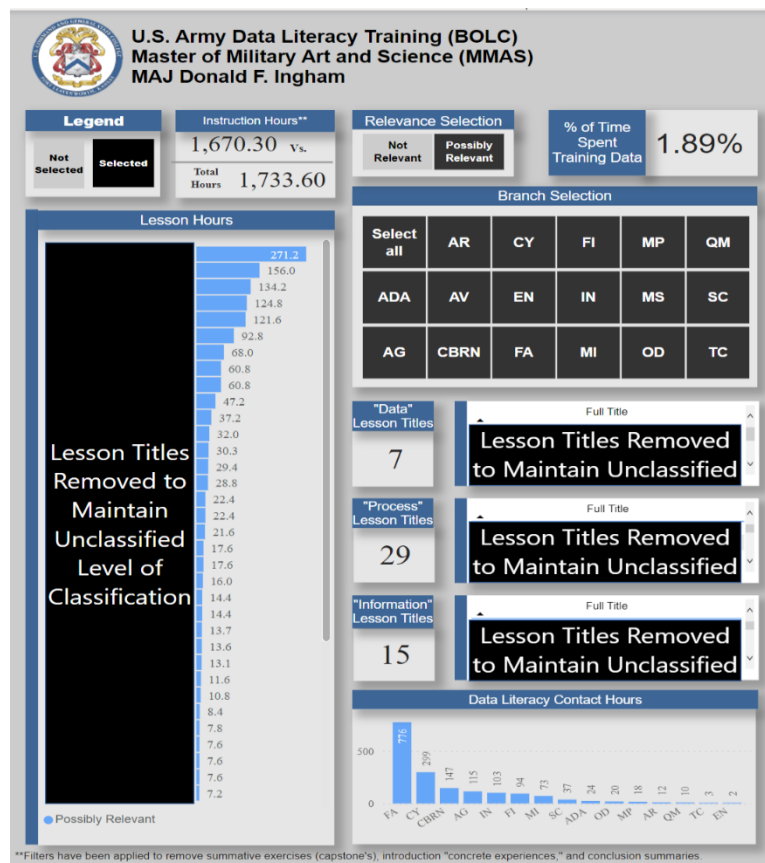


Figure 24. POI Quantitative Analysis

Source: Created by author.

After discounting the quantitative approach, branch names were hidden and each of the 1,318 distinct lesson titles were individually assessed to determine if the coursework is related to data literacy. This qualitative approach resulted in identifying 43 lesson titles, across eight branches, that related to data literacy. Many of these lessons were overlooked by the quantitative approach. Maintaining the coursework filter, this analysis determines that the average Army Officer is trained on data literacy relevant coursework during 3.6% of their time at BOLC (Figure 25). Branch acronyms are listed in Appendix H.

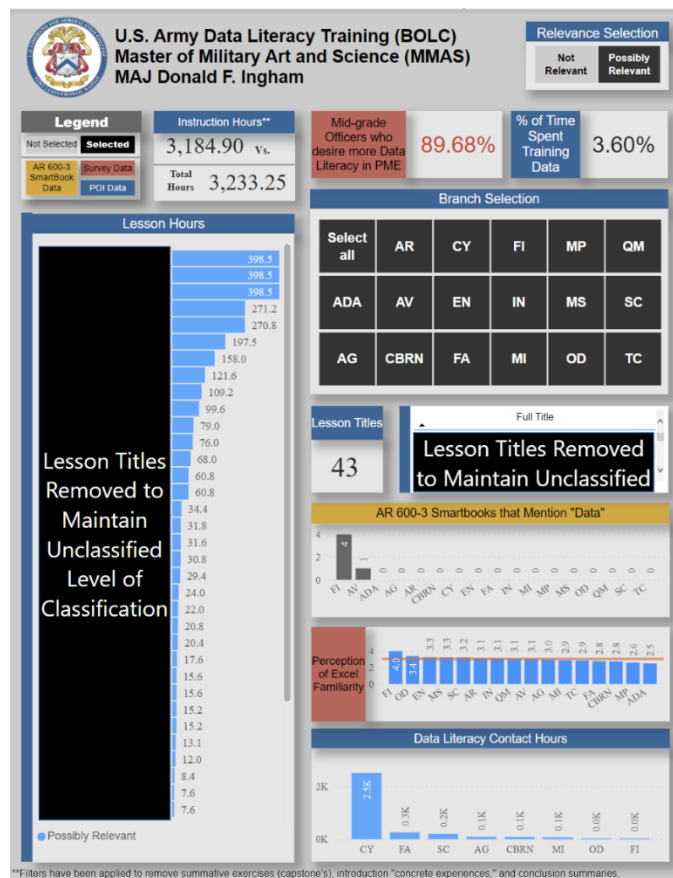


Figure 25. Initial POI Qualitative Analysis

Source: Created by author.

Upon further inspection, this method of content identification was also skewed. For example, the newly established Cyber Corps teaches an astounding 2,494.2 hours of data literacy relevant coursework, which immediately affects these numbers (Figure 26). Branch acronyms are listed in Appendix H. For clarity, total instruction hours include some content that is taught concurrently, especially during field time which is not restricted to a typical eight-hour training day. Furthermore, as Cyber is a new branch, no current CGSOC student had the opportunity to attend Cyber BOLC. Removing Cyber from our analysis tells a different story (Figure 27).

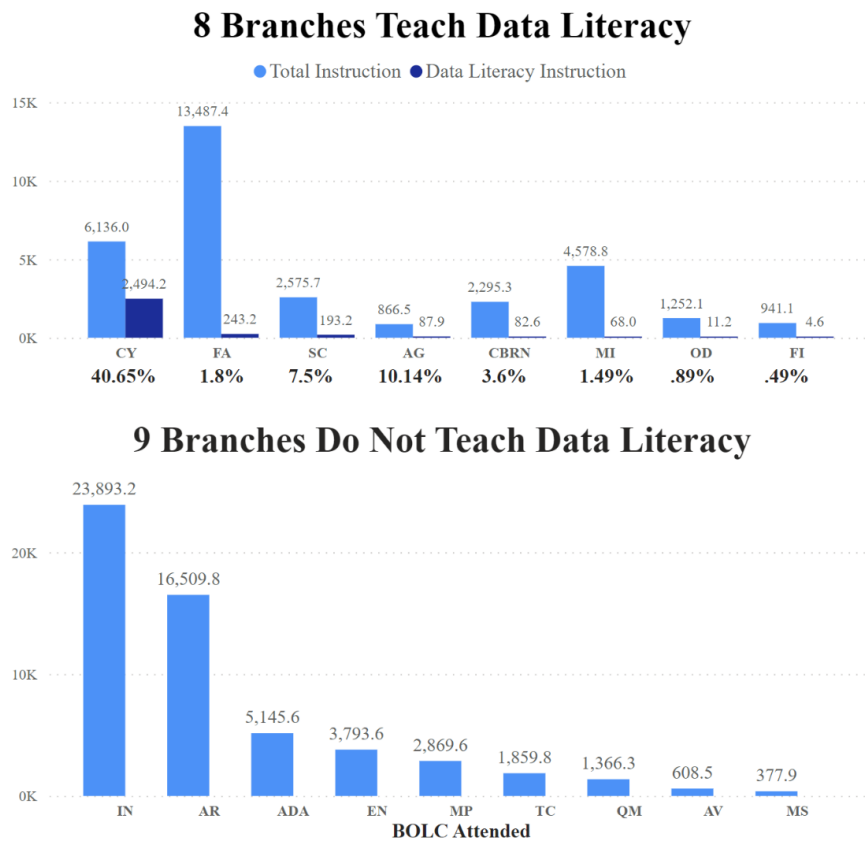


Figure 26. POI Qualitative Analysis – Branch Comparison

Source: Created by author.

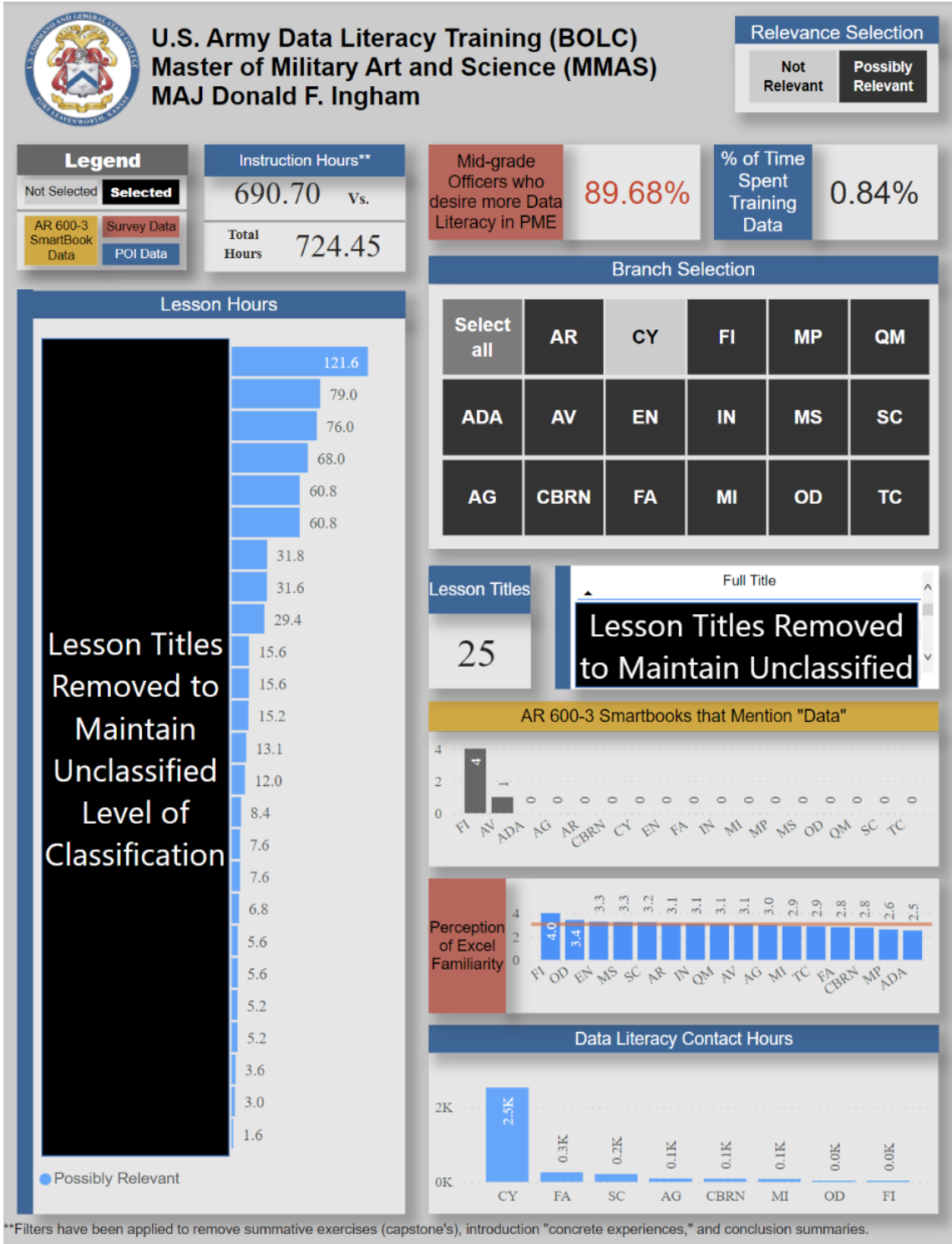


Figure 27. POI Qualitative Analysis – Removing Cyber

Source: Created by author.

After removing Cyber coursework, the number of data literacy relevant lesson titles drops to 25, and the average time an Officer spends training on this content falls to less than 1% (0.84%) of total BOLC training time. This includes nine branches that teach zero hours of data literacy: Infantry, Armor, Air Defense Artillery, Engineers, Military Police, Transportation Corps, Quartermaster, Aviation, and Medical Services. Of note, Aviation was one of the two branches that used the term “data” in the DA PAM 600-3 Smartbook, yet no data literacy relevant lesson title was found within Aviation BOLC. In summary, the 690.7 (non-Cyber) data literacy relevant course hours were found in only seven branches: Field Artillery, Signal Corps, Adjutant General, Chemical, Military Intelligence, Ordinance, and Finance.

Primary Research Question Findings

Upon conclusion of the exploratory analysis, the researcher addressed the primary research question: Is there a significant difference in the average amount of time mid-grade U.S. Army Officer’s spend manipulating data or analyzing information comparing across BOLC attended or current branch? To answer this question, survey data was input into Statistical Product and Service Solution (SPSS) software, which ran two one-way MANOVAs. The first one-way MANOVA used the independent variable BOLC attended, compared across two dependent variables of time spent manipulating data, and time spent analyzing information. This test returned a Wilks’ Lambda significance value of 0.009. As values below 0.05 indicate a statistically significant result, this number implies that there appears to be a significant difference across all dependent values as a function of the independent value. Once verified, the researcher tested the first set of hypotheses:

H1: The average amount of time a mid-grade Officer spends manipulating data or analyzing information is different when comparing PME BOLC attended.

H2: The average amount of time a mid-grade Officer spends manipulating data is different when comparing PME BOLC attended.

H3: The average amount of time a mid-grade Officer spends analyzing information is different when comparing PME BOLC attended.

H0: The average amount of time a mid-grade Officer spends manipulating data or analyzing information is the same regardless of PME BOLC attended.

SPSS returned a 0.071 significance value when comparing differences in time spent manipulating data across BOLC attended. As this number is slightly over the 0.05 indicator, this researcher has identified this as an interesting, yet inconclusive, data point. More research or a larger sample is necessary to reduce this number below the 0.05 standard. Conversely, SPSS returned a 0.027 significance value when comparing time spent analyzing information across BOLC attended. This value indicates statistical significance. Given these results, this researcher has evidence to support H3 with interesting, yet inconclusive, data to support H1. This analysis can be visualized in Figure 28.

Step 1)

Effect		Value	F	Hypothesis df	Error df	Sig.
BOLCAttended	Pillai's Trace	.227	1.708	34.000	454.000	.009
	Wilks' Lambda	.786	1.701 ^b	34.000	452.000	.009
	Hotelling's Trace	.256	1.694	34.000	450.000	.010
	Roy's Largest Root	.140	1.865 ^c	17.000	227.000	.022

Step 2)

Source	Dependent Variable	Sig.
BOLCAttended	Time Spent Manipulating Data	.071
	Time Spent Analyzing Information	.027

Key
Significant
Interesting
Does Not Correlate

Step 3)

Hypotheses:

- H1: The average amount of time mid-grade Officers spend **manipulating data** or **analyzing information** is different when comparing PME **BOLC attended**.
- H2: The average amount of time mid-grade Officers spend **manipulating data** is different when comparing PME **BOLC attended**.
- H3: The average amount of time mid-grade Officers spend **analyzing information** is different when comparing PME **BOLC attended**.
- H0: The average amount of time mid-grade Officers spend **manipulating data** or **analyzing information** is the same regardless of PME **BOLC attended**.

Figure 28. SPSS Results – Hypotheses H1-H3

Source: Created by author.

The second one-way MANOVA used the independent variable current branch, compared across two dependent variables of time spent manipulating data, and time spent analyzing information. This test returned a Wilks' Lambda significance value of 0.006. As values below 0.05 indicate a statistically significant result, this number implies that there appears to be a significant difference across all dependent values as a function of the independent value. Once verified, the researcher tested the second set of hypotheses:

H4: The average amount of time a mid-grade Officer spends manipulating data or analyzing information is different when comparing their current branch.

H5: The average amount of time a mid-grade Officer spends manipulating data is different when comparing their current branch.

H6: The average amount of time a mid-grade Officer spends analyzing information is different when comparing their current branch.

H0: The average amount of time a mid-grade Officer spends manipulating data or analyzing information is the same regardless of current branch.

SPSS returned a 0.023 significance value when comparing differences in time spent manipulating data across current branch. As this number is under the 0.05 indicator, this result indicates significance. Conversely, SPSS returned a 0.226 significance value when comparing time spent analyzing information across current branch. This value indicates that there is not enough evidence to suggest effects between current branch and time spent analyzing information. Given these results, this researcher has evidence to support H5. This analysis can be visualized in Figure 29.

Step 1)

Multivariate Tests^a

Effect		Value	F	Hypothesis df	Error df	Sig.
CurrentBranch	Pillai's Trace	.331	1.596	54.000	434.000	.007
	Wilks' Lambda	.695	1.597 ^b	54.000	432.000	.006
	Hotelling's Trace	.401	1.598	54.000	430.000	.006
	Roy's Largest Root	.251	2.015 ^c	27.000	217.000	.003

Step 2)

Tests of Between-Subjects Effects

Source	Dependent Variable	Sig.
CurrentBranch	Time Spent Manipulating Data	.023
	Time Spend Analyzing Information	.226

Key

- Significant
- Interesting
- Does Not Correlate

Step 3)

Hypotheses:

H4: The average amount of time mid-grade Officers spend **manipulating data** or **analyzing information** is different when comparing their **current branch**.

H5: The average amount of time mid-grade Officer spend **manipulating data** is different when comparing their **current branch**.

H6: The average amount of time mid-grade Officers spend **analyzing information** is different when comparing their **current branch**.

H0: The average amount of time mid-grade Officers spend **manipulating data** or **analyzing information** is the same regardless of **current branch**.

Figure 29. SPSS Results – Hypotheses H4-H6

Source: Created by author.

Secondary Research Question 1 Findings

After confirming the primary research question, the researcher moved to answer the first secondary research question: How much does Professional Military Education data literacy training change the amount of time mid-grade U.S. Army Officers spend manipulating data? To conduct this bivariate correlation analysis, the researcher created a scatterplot in Excel that uses the POI generated training hours on the X-axis and the averaged survey responses for data manipulation time on the Y-axis, organized by BOLC attended. The initial chart did not provide significant insights – until the branches with zero data literacy in PME were filtered out. Upon applying this filter, a relevant result

became apparent. This chart displays a strong negative correlation, with a R^2 value of 0.6093. As this value is greater than 0.5, the regression line is a good fit for the datapoints. The regression line equation is $y = -0.0049x + 3.4535$. Using algebra, this research can now identify a likely number of PME data literacy training hours that can reduce the average number of hours spent manipulating data. For example, suppose the Army wants to reduce the average daily amount of time Officers spend manipulating data by 18 minutes (the difference between the time an average Officer spends manipulating data versus analyzing information). Using the regression line, this research determines that an addition of 61.22 hours of data literacy relevant BOLC instruction will likely result in this 18-minute reduction of daily manipulation time (Figure 30). Branch acronyms are listed in Appendix H.

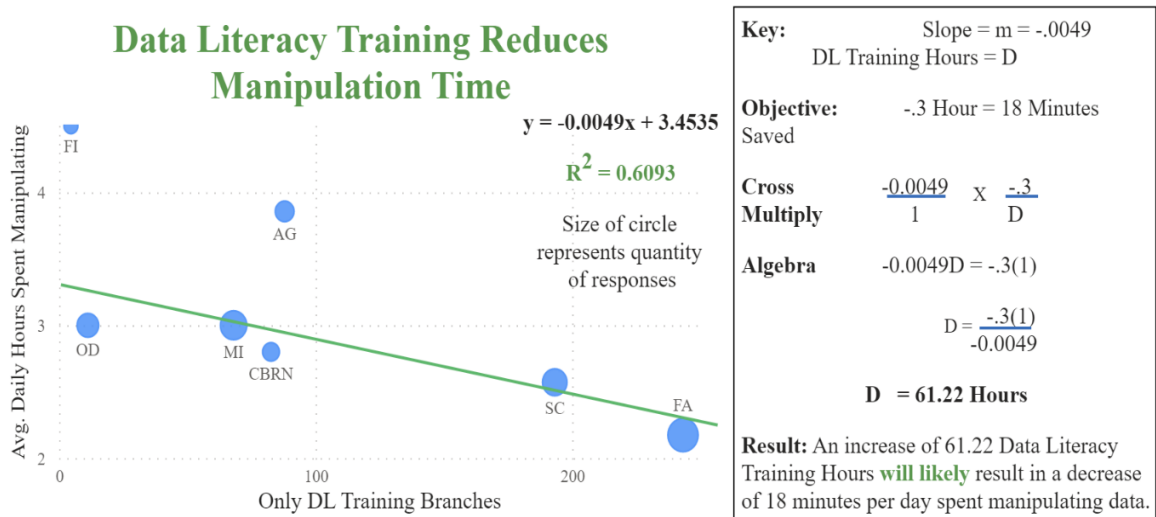


Figure 30. Data Literacy Training Reduces Manipulation Time

Source: Created by author.

Secondary Research Question 2 Findings

After confirming the first secondary research question, the researcher moved to answer the second secondary research question: How much does Professional Military Education data literacy training change the amount of time mid-grade U.S. Army Officers spend analyzing information? To conduct this bivariate correlation analysis, the researcher again created a scatterplot in Excel that uses the POI generated training hours on the X-axis and the averaged survey responses for analysis time on the Y axis, organized by BOLC attended. The initial chart did not provide significant insights – until the branches with zero data literacy in PME were filtered out. Upon applying this filter, an interesting result became apparent. This chart displays a negative correlation, with a R^2 value of 0.2861. As this value is close to (but still below) 0.5, this researcher has identified this regression line as a reasonable fit, which may improve with more data. The resulting regression line equation is $y = -0.0041x + 3.132$. Using algebra, this research can now identify potential impacts of adding PME data literacy training hours on the amount of daily time spent analyzing information. For example, if the Army implements the previous recommendation and adds 61.22 hours of data literacy training (saving 18 minutes of daily manipulation time), this may additionally result in the reduction of daily analysis time by 15 minutes (Figure 31). Branch acronyms are listed in Appendix H.

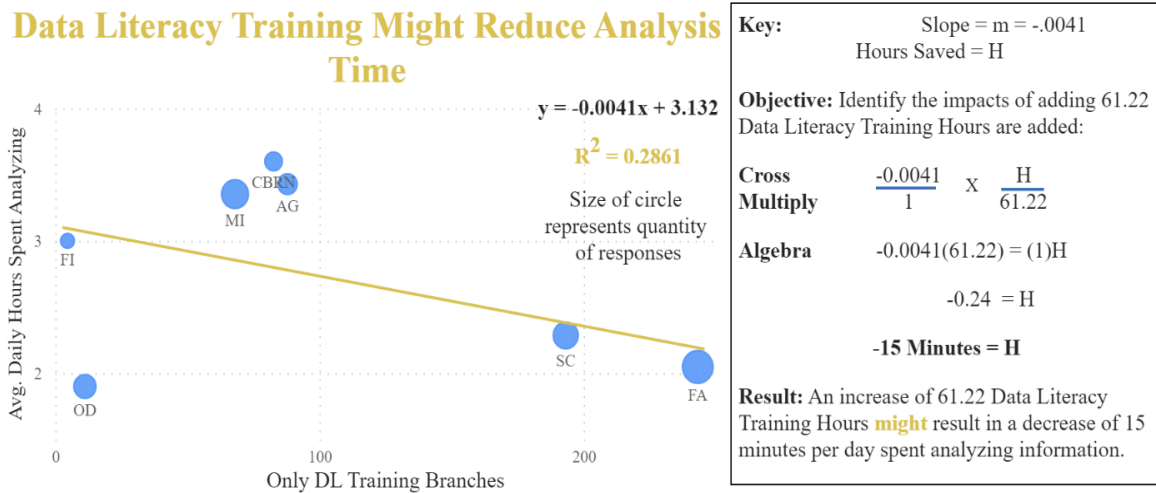


Figure 31. Data Literacy Training Might Reduce Analysis Time

Source: Created by author.

Initially, this research assumed that the results for secondary research question two would indicate a positive correlation, instead of the negative correlation that the data implies. The original assumption was that, as there are a finite number of hours each day, fewer hours spent manipulating data would allow Soldiers to reallocate that time to information analysis. Upon further reflection, a negative correlation makes sense as well. Presumably, when Soldiers become more data literate, their ability to quickly analyze information also increases. This data indicates that these two skills are similarly affected by the amount of time in DL training – indicated by similar slope values (-0.0049 slope for manipulation time, versus -0.0041 slope for analysis time). This result implies that the net benefit to each individual Officer (amount of time saved per day) is cumulative. For example, this finding implies that if 61.22 hours of PME data literacy training are added, the result would be 18 minutes of time saved per day during data manipulation and additionally 15 minutes of time saved per day during information analysis, resulting in a

total of 33 minutes of time saved each day. Assuming 260 work days per year, this calculates to 8,580 minutes (143 hours) saved per year – per U.S. Army Officer.

Secondary Research Question 3 Findings

After finding evidence to support the second secondary research question, the researcher moved to answer the third secondary research question: How much does Professional Military Education data literacy training change a mid-grade U.S. Army Officer's familiarity with Excel? To conduct this bivariate correlation analysis, the researcher again created a scatterplot in Excel that uses the POI generated training hours on the X-axis, and the averaged survey responses for Excel familiarity on the Y-axis, organized by BOLC attended. Again, the initial chart did not provide significant insights until the branches with zero data literacy in PME were filtered out. Upon applying this filter, an interesting result became apparent. This chart displays a negative correlation, with a R^2 value of 0.3379. As this value is close to (but still below) 0.5, this researcher has identified this regression line as a reasonable fit, which may improve with more data. The resulting regression line equation is $y = -0.0015x + 3.2156$. Using algebra, this research can now identify potential impacts of adding PME data literacy training hours on the amount of perceived Excel familiarity. For example, if the Army implements the previous recommendation and adds 61.22 hours of data literacy training, this may additionally result in the reduction of perceived Excel familiarity by 0.09 of a point – on a five-point scale (Figure 32). Branch acronyms are listed in Appendix H.

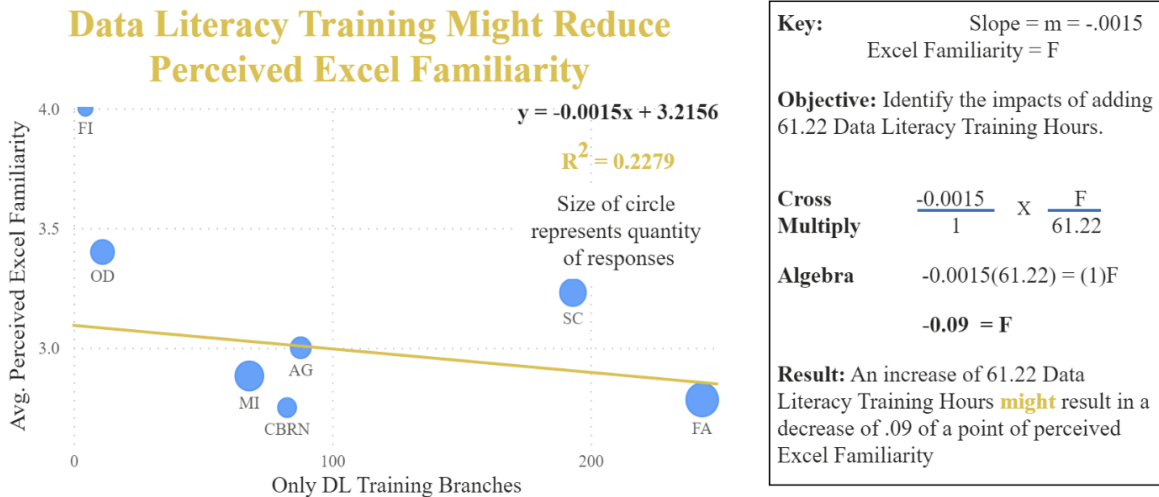


Figure 32. Data Literacy Training Might Reduce Perceived Excel Familiarity

Source: Created by author.

Initially, this researcher assumed that the results for secondary research question three would indicate a positive correlation instead of the negative correlation that the data implies. The original assumption was that, as data literacy training increases, Soldiers would feel more comfortable using Excel formulas in their daily routines. Upon further reflection, a negative correlation might have presented itself due to the Dunning-Kruger effect, which is a cognitive bias that indicates that people initially overestimate their skills before reducing their estimation after they become more competent in a given subject.¹ Presumably, when Soldiers become more data literate, their confidence (and thus their perception of their Excel familiarity) decreases instead of increases.

Secondary Research Question 4 Findings

After finding evidence to support the third secondary research question, the researcher moved to answer the fourth and final secondary research question: How much

does Professional Military Education data literacy training change a mid-grade U.S. Army Officer's perception of their own data literacy? To conduct this bivariate correlation analysis, the researcher again created a scatterplot in Excel that uses the POI generated training hours on the X-axis, and the averaged survey responses of data literacy perception on the Y-axis, organized by BOLC attended. Again, the initial chart did not provide significant insights. Furthermore, no interesting results were found even after branches with zero data literacy in PME were filtered out. This chart displays no strong correlation and a very weak R^2 value. (Figure 33) Branch acronyms are listed in Appendix H.

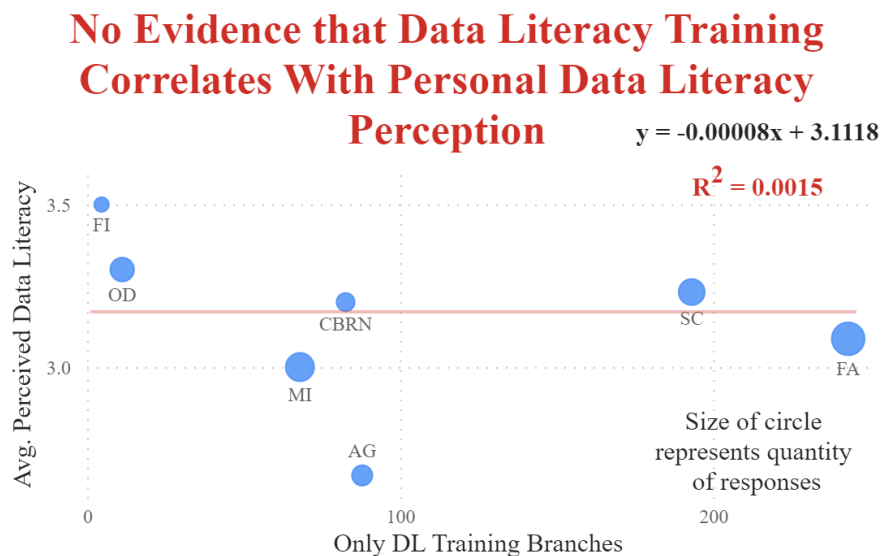


Figure 33. No Evidence to Support Correlation between Data Literacy Training and Perception

Source: Created by author.

Initially, this researcher assumed that the results for secondary research question four would indicate a strong positive correlation. The original assumption was that, as data literacy training increases, Soldiers would feel more comfortable with their own data literacy. As with secondary research question three, these results may have presented due to the Dunning-Kruger effect.² An additional consideration may be that this research was not able to find any BOLC PME lesson titles that directly address data literacy. All lesson titles that have been identified as data literacy relevant are only tangentially related to data literacy (for example, a lesson title of “programming” would be identified as data literacy relevant, yet this content is not directly training “data literacy”). Assessing students who were directly trained on data literacy may result in a stronger correlation. Given these results, the researcher is not able to make any further conclusions in regards to research question four.

Summary

This research determined three initially interesting findings. First, 89.68% of mid-grade U.S. Army Officers believe that there is not enough data literacy training in PME. Second, the average mid-grade U.S. Army Officer spends 2.78 hours per day manipulating data, versus 2.48 hours per day analyzing information. Combined, this is a significant portion of any workday and interestingly equates to 18 more minutes per day spent manipulating data, rather than analyzing information. Third, this research qualitatively determined that, after removing Cyber coursework, only 0.84% of BOLC PME time is spent training data literacy relevant coursework – a total of 690.7 hours of total content spread between eight branches, with nine branches receiving zero data literacy relevant coursework.

This research found two significant findings. First, SPSS confirmed the primary research question through statistical evidence for Hypothesis 3 and Hypothesis 5, with interesting but inconclusive evidence for Hypothesis 1. Given this, the researcher can state with confidence the following three assertions. First, there is a significant difference in the average amount of time mid-grade U.S. Army Officers spend manipulating data compared by their current branch. Second, there is a significant difference in the average amount of time mid-grade U.S. Army Officers spend analyzing information compared by the Basic Officer Leadership Course (BOLC) they attended. Finally, there is statistical evidence that may indicate that the average amount of time mid-grade U.S. Army Officers spend both manipulating data and analyzing information is significant when compared across BOLC attended.

The second significant finding confirmed the first secondary research question, allowing the researcher to state that data literacy training very likely reduces the average amount of daily time spent manipulating data. Approximately 3.5 hours of increased data literacy training likely reduces the amount of daily time spent manipulating data by one minute. Calculating this conclusion out, if the Army decided to reduce the amount of daily time spent manipulating data by 18 minutes per day (the difference between the reported average amount of time an Officer spends manipulating data and the amount of time they spend analyzing information) it would require an addition of 61.22 hours of BOLC training. This additional coursework will likely save 4,680 minutes (78 hours) of annual work time – per Officer.

This research found an additional two unexpected and interesting, albeit inconclusive results for secondary research questions three and four. First, contrary to

original assumptions, data literacy training might reduce analysis time. Approximately 4.1 hours of increased data literacy training might reduce the amount of daily time spent analyzing information by one minute. If these results hold, an addition of the aforementioned 61.22 hours of BOLC training will reduce the average daily time Officers spend analyzing information by an additional 15 minutes. Combining this 15-minute reduction with the previously saved 18 minutes of manipulation time equates to saving a total of 33 minutes per day (8,580 minutes/143 hours per year). The second unexpected finding also contradicts original assumptions in that data literacy training might actually reduce perceived Excel familiarity, although only slightly. According to this data, an increase of over 650 hours of data literacy training might reduce the amount of perceived Excel familiarity by one point (for example, from “very familiar” to “familiar,” or from “familiar” to “unfamiliar.”) For context, an increase of the suggested 61.22 data literacy training hours might reduce Officers perception of their own Excel familiarity by 0.09 of a point – on a five-point scale.

Finally, there was one inconclusive result for secondary research question four. This research could not find any evidence to support research question four and is thus unable to determine how much data literacy training may correlate with an individual’s perception of their own data literacy. This result (as well as the previous result articulating Excel familiarity) may stem from difficulties in quantitatively measuring perception, or the results may be an example of the Dunning-Kruger effect.³

¹ Brian Duignan. “Dunning-Kruger effect.” Encyclopedia Britannica, September 8, 2020. <https://www.britannica.com/science/Dunning-Kruger-effect>.

² Ibid.

³ Ibid.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

Introduction

The intent of this section is to elevate the preceding predictive analysis, into a prescriptive analysis, by suggesting institutional changes that will improve the force-wide understanding of data literacy. This analytical advancement is possible because of the statistical success found while exploring the research questions. In summary:

1. Primary Research Question: Is there a significant difference in the average amount of time mid-grade U.S. Army Officer's spend manipulating data or analyzing information comparing across BOLC attended or current branch?

This answer is best described in three parts:

1. There *is a significant difference* in the average amount of time mid-grade U.S. Army Officers spend *manipulating data* compared by their *current branch*.
 2. There *is a significant difference* in the average amount of time mid-grade U.S. Army Officers spend *analyzing information* compared by the *Basic Officer Leadership Course (BOLC) they attended*.
 3. There *is statistical evidence that may indicate* that the average amount of time mid-grade U.S. Army Officers spend *both manipulating data and analyzing information* is significantly different when compared across *BOLC attended*.
2. Secondary Research Question 1: How much does Professional Military Education data literacy training change the amount of time mid-grade U.S. Army Officers spend manipulating data?

1. Answer confirmed: Data literacy training *reduces* manipulation time.

2. How much: Approximately *3.5 hours of increased* data literacy training *likely reduces* the amount of daily time spent manipulating data *by one minute*.
3. Secondary Research Question 2: How much does Professional Military Education data literacy training change the amount of time mid-grade U.S. Army Officers spend analyzing information?
 1. Evidence to support: Data literacy training *might reduce* analysis time.
 2. How much: Approximately *4.1 hours of increased* data literacy training *might reduce* the amount of daily time spent analyzing information *by one minute*.
4. Secondary Research Question 3: How much does Professional Military Education data literacy training change a mid-grade U.S. Army Officer's familiarity with Excel?
 1. Evidence to support: Data literacy training *might reduce* an Officers perception of their own Excel familiarity.
 2. How much: Approximately *660 hours of increased* data literacy training *might reduce* the amount of perceived Excel familiarity *by one point* (on a five-point scale).
5. Secondary Research Question 4: How much does Professional Military Education data literacy training change a mid-grade U.S. Army Officer's perception of their own data literacy?
 1. No Evidence: This research was *unable to determine* how much (or if) data literacy training correlates with an individual's perception of their own data literacy.

Implications

This evidence has one key implication: a relatively simple policy change can affect the input (institutional training), which will drive a significant and calculable effect on the output (time spent manipulating data/analyzing information). The potential army-wide time savings are enormous. As an example, 61.22 additional hours of data literacy training likely results in 18 minutes saved during data manipulation and an additional 15 potential minutes saved during information analysis, totaling 33 minutes per-Officer per-day (Figure 34). Assuming a total Officer population of 79,310 (Fiscal Year 2022), this equates to a staggering *43,620.5 Army work hours saved – every day!*¹ Even more astounding, this equates to *11,341,330 annual Army work hours saved!* (This calculation assumes 260 work days in a year) What could the Army accomplish with 11,341,330 annual work hours?

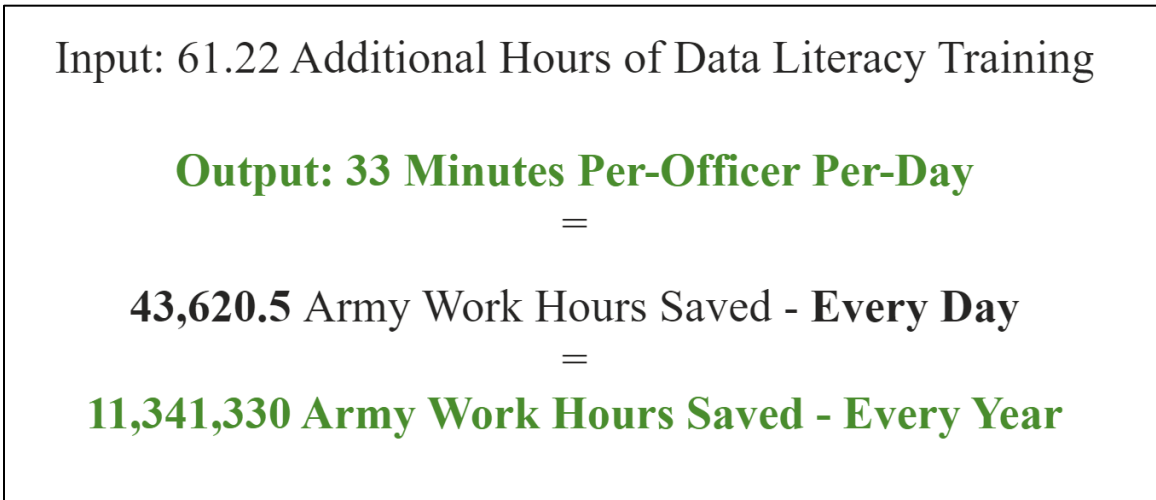


Figure 34. Potential Impact of Data Literacy Training on Time

Source: Created by author.

Three Recommendations

The first recommendation is a simple yet fundamental doctrine update. Incorporate the concept – and definition – of data literacy into ADP 6-0 and FM 6-0. The suggested definition is that “data literacy is the ability to read, work with, analyze and communicate with data.”² This update should clarify that the “process” for turning data into information is a human endeavor, not an entirely automated one. All Officers are expected to be data literate in the same way that all cooks are expected to be able to use raw ingredients.

The second recommendation relies on the fact that the process used to turn data into information is not branch-specific. As such, data literacy training should not be branch specific. Fortunately, the Army has a process for standardizing Officer initial entry training called “common-core.” This researcher recommends that the Army mandate a 62-hour course load of data literacy training in BOLC common-core. This will ensure a standardized curriculum and universal instruction. As a side note, none of the lessons identified in this analysis directly taught data literacy. As such, time-saving increases may be even more robust if concentrated data literacy training is delivered instead of merely data literacy relevant instruction.

The third recommendation refers to curriculum development. This researcher recommends that the Army organize course curriculum to reflect the standards articulated in the definition of data literacy. Each aspect of the definition should drive specific and actionable learning objectives. This standard will align doctrine with instruction to create a consistent understanding of data. An example of how this might look, with recommended lesson subjects, can be found in Figure 35.

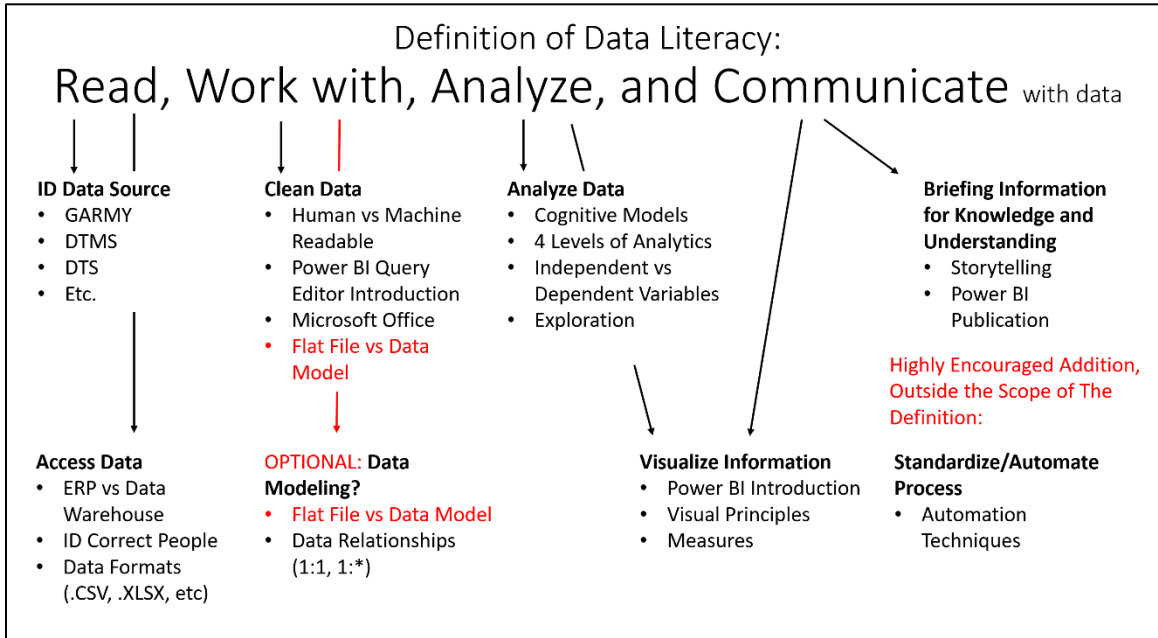


Figure 35. Recommended Data Literacy Learning Objectives

Source: Created by author.

Recommendations for Future Research

The primary recommendation for future research is to avoid quantitative survey questions that measure a respondent’s perception of their abilities. Secondary research questions three and four are significantly more difficult to analyze, which have arisen due to cognitive biases that were not considered while designing the survey questions. Furthermore, these questions provide far less actionable information than secondary research questions one and two.

The second recommendation is to use this methodology to expand the research to include Captains Career Course (CCC) data literacy relevant content. This inclusion would (presumably) drastically increase the depth and value of the final dataset. With this inclusion, a significantly higher number of survey respondents would have a unique total

of data literacy training hours. To illustrate the problem, as the data currently reads, two Officers who attended Infantry BOLC are each calculated at zero data literacy training hours – even if one of these individuals attended the Cyber Captains Career Course and was thus inundated with data literacy relevant instruction. If someone is interested in undertaking this work, this researcher would provide any base documents or reference material.

The third recommendation is to investigate the effects of data literacy on decision-making. For example, present survey respondents with two different products derived from the same dataset – one produced by a data literate individual and represents truly visualized information, and the other produced by someone with low data literacy and represents merely processed data. Compare both the quality of insights that survey respondents made, as well as the time it took respondents to answer each question. This researcher assumes that truly visualized information will generate both better and faster insights.

Conclusion

Time is an increasingly scarce resource. Every minute spent manipulating data is time that cannot otherwise be used in more productive ways. As data becomes increasingly ubiquitous within the modern operating environment, basic data literacy becomes imperative. Without these skills, Officers must either compensate by increasing the amount of time they commit to data manipulation, or sacrifice a clear understanding of their operating environment. Fortunately, there is a direct correlation between training hours and time saved.

Despite inconsistent execution, some Civilian examples of this type of curriculum do exist. For example, Georgetown University in Qatar designed a one-credit data literacy class. Operating under a slightly different definition of data literacy (“the ability to find, analyze, interpret and effectively communicate data and the key insights derived from it.”³) this institution developed six lessons. Lessons are entitled: “data visualization and basic design principles,” “data discovery and processing: how to find, treat and cite data,” “charts, graphs, and maps. Using Venngage software,” “storytelling with data,” “tools for independent learning and skill building,” “final projects presentation. Wrap-up and course assessment.”⁴ An example of a student submitted final project can be seen in Figure 36.



Figure 36. Data Literacy Class Final Project/Visualized Information

Source: Tatiana Usova and Robert Laws, “Teaching a One-Credit Course on Data Literacy and Data Visualisation,” *Journal of Information Literacy* 15, no. 1 (January 2021): 90, <https://www.ariadne.ac.uk/JIL/article/view/PRJ-V15-I1-3>.

As a final note, it is interesting to compare the current U.S. Army modernization effort with recent events within the Russian military. Some accounts are showing that, despite massive equipment investments, many of Russia’s most lethal and modern assets are not being employed in the current war with Ukraine due to a lack of training. Senior leaders appear unwilling to risk placing these investments under the control of Service Members who have very little experience operating the equipment.⁵ Training can be expensive, difficult to quantify, and prone to atrophy, yet is fundamentally important. Soldiers win wars; equipment only helps if Soldiers know how to employ it. As the U.S. Army undertakes its own data-centric modernization effort, it is important that the best financed military in the world does not similarly forget that humans require training, and war is a human endeavor. Learning a skill like data literacy is no different than learning how to cook; education of any vein requires an investment, practice, and feedback. Now that we have the recipe, what remains to be seen is if the Army will follow it.

¹ Erin Duffin. “Total military personnel of the U.S. Army, for fiscal years 2020 to 2022, by rank.” Statista, June 21, 2021, <https://www.statista.com/statistics/239383/total-military-personnel-of-the-us-army-by-grade/>.

² Morrow, *Be Data Literate*, 36.

³ *Ibid.*, 84.

⁴ Usova and Laws, “Teaching a One-Credit Course on Data Literacy and Data Visualisation,” 90.

⁵ Justin Bronk, “The Mysterious Case of the Missing Russian Air Force,” RUSI, February 28, 2022. <https://rusi.org/explore-our-research/publications/commentary/mysterious-case-missing-russian-air-force>.

GLOSSARY

- Analyzing Information. This author uses this term to mean: exploring the results of manipulated data (definition included in the survey).
- Artificial Intelligence. “within predictive and prescriptive analytics, artificial intelligence can play a crucial role as the power of computing and processing speed can bring about great predictions and help us [humans] understand what we should do.”¹
- Business Intelligence. “tools and sourcing of data that organizations can use to succeed with the four levels of analytics.”²
- Machine Learning. “the field where algorithms are learning by themselves, and can improve for the betterment of our organization, data, analytics, etc.”³
- Algorithm. “step-by-step procedure for solving a problem or accomplishing some end.”⁴
- Raw Data. “raw data is the initial state of data you have collected, receive or downloaded that has not yet been subjected to any statistical or transforming treatment.”⁵
- Data (as defined in doctrine). “in the context of decision making, *data* consists of unprocessed observations detected by a collector of any kind (human, mechanical, or electronic)”⁶
- Data Source. “Origins of raw data.”⁷
- Data Visualization. “the visual representation and presentation of data to facilitate understanding.”⁸
- Data Literacy. “data literacy is the ability to read, work with, analyze, and communicate with data.”⁹
- Descriptive Analytics. “building of reports, dashboards, and observations that help an organization know what has happened to that organization in the past, or what is currently happening.”¹⁰
- Diagnostic Analytics. “getting the insight in the data, learning the drivers and why things happened.”¹¹
- Predictive Analytics. saying “it will happen or [taking] the time to estimate something that will happen” in the future.¹²
- Prescriptive Analytics. “prescribing what should be done what the data and analytics, and what . . . decision should be made.”¹³

Processed (as defined in doctrine). “processing information describes the act of cataloging data via established and usually routine sets of procedures to link or fuse it with other related data in order to create information.”¹⁴

Manipulating Data. This author uses this term to mean: exporting online databases, changing excel spreadsheets, transferring data into PowerPoint, etc. (definition included in the survey).

Mid-grade. This author uses this term to mean: Senior Captain or junior Major.

Human Readable. This research uses the term “human readable” to refer to data that has been transformed from “raw data” into a standard matrix that humans can interact with directly.

Information (as defined in doctrine). “in the context of decision making, *information* is data that has been organized and processed in order to provide context for further analysis.”¹⁵

Processed Data. This research uses the term “processed data” to refer to products that have analyzed data, but do not elevate to the level of useful information.

¹ Morrow, *Be Data Literate*, 187.

² Ibid., 183.

³ Ibid., 189.

⁴ Ibid., 188.

⁵ Andy Kirk, *Data Visualization* (Los Angeles: Sage 2019), 11.

⁶ HQDA, ADP 6-0, 2-4.

⁷ Kirk, *Data Visualisation*, 11.

⁸ Ibid., 15.

⁹ Morrow, *Be Data Literate*, 36.

¹⁰ Ibid., 21.

¹¹ Ibid., 24.

¹² Ibid., 27.

¹³ Ibid., 29.

¹⁴ HQDA, FM 6-0, 3-9.

¹⁵ HQDA, ADP 6-0, 2-4.

APPENDIX A:

SURVEY QUESTIONS AND INTENT

#	Question	Intent
1	<p>My current (or former) rank is:</p> <ul style="list-style-type: none"> • Civilian (never served) • Civilian (no longer serving) • Enlisted • 2LT • 1LT • CPT • MAJ • Lieutenant Colonel • Other (Fill In) 	<p>Independent variable, used to specifically identify the CPT/MAJ population.</p> <p>Enables the “mid-grade” portion of the research questions.</p>
2	<p>My current Branch is?</p> <ul style="list-style-type: none"> • Acquisitions • Adjutant General • Air Defense • Any Army Medical Department Branch • Armor • Aviation • Chaplain • Chemical • Civil Affairs • Cyber • Engineer • Field Artillery • Finance • Infantry • Judge Advocate General • Logistics • Military Intelligence • Military Police • Ordinance • Psychological Operations • Quartermaster • Signal • Special Forces • Transportation • FA ____ (Fill In Number) • Other (Fill In) 	<p>Independent variable, used to explore potential correlation between current branch and time spent manipulating or analyzing data.</p> <p>Enables the “current branch” portion of the primary research question.</p>

3	<p>What branch Basic Officer Leaders Course (or historical equivalent) did you attend?</p> <ul style="list-style-type: none"> • Adjutant General • Air Defense • Armor • Army Medical Department • Aviation • Chemical • Cyber • Engineer • Field Artillery • Finance • Infantry • Military Intelligence • Military Police • Ordinance Corps • Quartermaster Corps • Signal • Transportation Corps • Other (Fill In) 	<p>Independent variable, used as the primary key to align with BOLC data.</p> <p>Enables the “BOLC attended” portion of the primary research question, and the “Data Literacy training” portion of the secondary research questions.</p>
4	<p>What Captains Career Course (CCC) did you attend?</p> <ul style="list-style-type: none"> • Adjutant General • Air Defense Artillery • Army Medical Department • Aviation • Chemical, Biological, Radiological, Nuclear, and Explosive • Combined Logistics • Cyber • Engineer • Field Artillery • Finance • Maneuver • Military Intelligence • Military Police • Signal • Sister Service CCC Equivalent • Other (Fill In) 	<p>Independent variable, used to expand research into CCC.</p> <p>Enables the “Data Literacy training” portion of the secondary research question.</p>
5	<p>Did you attend Reserve Component CCC</p> <ul style="list-style-type: none"> • Yes • No 	<p>Independent variable, used to eliminate RCCC to ensure equivalence if research is expanded.</p>

6	<p>On an average Army day (while not in student status), how much time did you spend <u>manipulating</u> data? (Exporting online databases, changing excel spreadsheets, transferring data into PowerPoint, etc.)</p> <ul style="list-style-type: none"> • None • 1 hour • 2 hours • 3 hours • 4 hours • 5 hours • 6 hours • 7 hours • Over 7 hours 	<p>Dependent variable, used to identify wasted or inefficient time.</p> <p>Enables the “manipulating” portion of the primary research question and first secondary research question.</p>
7	<p>On an average Army day (while not in student status), how much time did you spend <u>analyzing</u> information? (Exploring the results of the manipulated data)</p> <ul style="list-style-type: none"> • None • 1 hour • 2 hours • 3 hours • 4 hours • 5 hours • 6 hours • 7 hours • Over 7 hours 	<p>Dependent variable, used to identify time spent analyzing information.</p> <p>Enables the “analyzing” portion of the primary research question and second secondary research question.</p>
8	<p>How familiar are you with excel formulas?</p> <ul style="list-style-type: none"> • Very Unfamiliar - I have never successfully created/used excel formulas • Unfamiliar - I have created/used simple excel formulas (ex. SUM) • Familiar - I have successfully created/used multiple excel formulas in the same document • Very Familiar - I am comfortable creating/using advanced excel formulas • Extremely Familiar - I create/use complicated excel formulas multiple times a week 	<p>Dependent variable, used as an indicator for personal perception of own data literacy.</p> <p>Enables the “Excel formula” portion of the third secondary research question.</p>

9	<p>On a scale of 1-5, how Data Literate do you perceive yourself to be? (Data Literacy is defined as your ability to “read, work with, analyze, and communicate with data”)</p> <p>Scale Definitions</p> <ul style="list-style-type: none"> • 1 (I find working with data confusing, and I avoid it when possible) • 2 (I find working with data difficult, and I am not efficient at processing data) • 3 (I find working with data reasonable, although I don’t consider myself an expert) • 4 (I find working with data simple, and I can confidently process data) • 5 (I find working with data very easy, and incorporate data processing into my daily routines) 	<p>Dependent variable, used to determine respondents’ personal opinion of own data literacy.</p> <p>Enables the “perception of own data literacy” portion of the fourth secondary research question.</p>
10	<p>Do you believe Professional Military Education (PME) has provided enough Data Literacy education/training (Data Literacy is defined as an ability to “read, work with, analyze and communicate with data”)?</p> <ul style="list-style-type: none"> • No • Yes 	<p>Dependent variable, used to determine respondents’ personal opinion of PME.</p> <p>Enables an overallpercentage.</p>

APPENDIX B:
SURVEY RECRUITMENT EMAIL

Hello from QAO,

This survey is part of a CGSOC MMAS study. Your experiences and inputs are appreciated. The information being gathered addresses Army training of data processing and your experience of data processing at your jobs. Your confidential responses will be provided to the researcher for analysis and publication. Please take **under 5 minutes** to complete this survey with your honest and candid responses to help the Army improve its educational institutions.

Please select the following link to take this survey:

//hyperlink//

This survey will be available from DATE through DATE

The researcher would like to sincerely thank you for your time and participation.

Please contact the undersigned if you have any difficulties accessing the survey

APPENDIX C:

SURVEY INFORMED CONSENT

A Recipe for Data Literacy: Modernizing the Human Domain Consent to Participate in a Research Study

This is a research study conducted in order to develop a baseline of data literacy knowledge within the Army, and is supporting MAJ Ingham's ongoing MMAS degree. This form provides information on your rights as a research participant in the above-named study and of the responsibilities the I have during this study. The Command and General Staff College (CGSC) has approved this study and supports the research.

Purpose of the Research Study

This research is designed to determine the baseline knowledge of data literacy among the U.S. Army population. It will be used to develop recommendations and guidance for developing or implementing data literacy education or training in professional military education institutions.

Procedures

Provide the following:

1. The researcher will invite 774 participants, targeting 258 respondents.
2. Each survey will take approximately five minutes.
3. Upon agreeing to take the survey, participants will be asked to take a 10-question survey.
4. Under no condition will the researcher terminate participation in this survey.
5. Nothing within this survey is considered classified, or could constitute a potential violation of the UCMJ or criminal law. Furthermore, no comments

will be elicited that could place an individual's clearance, credentials, or other privileged access or duty at risk.

Risks

DoD employees should consult with their commander or supervisor prior to participating in this study.

Participation in this study entails a risk level that has been determined is both very low likelihood and very low severity. Questions revolve around institutional training and perception of data, data literacy, and related fields.

Benefits

This is a research study and there is no expectation that you will receive any direct benefit from participation.

Compensation

Participants will not be compensated for their participation.

Confidentiality

Provide the following:

1. Confidentiality of records will be maintained. All identifying information will be removed by the Quality Assurance Office before the researcher has access to the data. Additionally, the Human Research Protection Program personnel or a DoD designee may inspect the records to ensure confidentiality.
2. Once collected, data will be maintained on a jump drive, which will be located in a secure fire and waterproof safe, behind a locked door, for no less than three years after the publication of this research.

3. All data obtained about you, as an individual, will be considered privileged and held in confidence; you will not be identified in any presentation of the results unless you wish so. Complete confidentiality cannot be promised to subjects, particularly to subjects who are military personnel, because information bearing on your health might be required to be reported to appropriate officials.
4. All data related to this study will remain secured for a period of not less than three years from the approval date for the research study.

Contacts for Additional Assistance

For answers to pertinent questions about the research, the research subjects' rights, and whom to contact in the event of a research-related injury, please contact:

1. Principal Investigator: MAJ Donald Ingham at donald.f.ingham.mil@army.mil.
2. CGSC Human Protections Director: Dr. Michelle A. Miller at michelle.a.miller48.civ@army.mil.

Voluntary Participation

Participation in a research study is voluntary. Anyone who is asked to be in a research study may say no. No one has to become a research subject. If you start a research study, you may stop at any time. You do not need to give a reason. No one can discriminate against you or treat you differently if you choose not to be in a research study or later decide to stop your participation.

Statement of Consent

By selecting the “Yes, I would like to participate” button, I am acknowledging that I have read this form and its contents were explained. I agree to be in this research study for the purposes listed above. All of my questions were answered to my satisfaction.

APPENDIX D:

POI DOCUMENT ANALYSIS

POIs Accessed in TDC as of DEC 2021			
Branch	POI Course Number	Version	Fiscal Year
ADA	2-44-C20B	3	2020
AG	7-12-C20B	8	2021
AR	2-17-C20B	3	2020
AV	2-1-C20B	3	2019
CBRN	4-3-C20B	13	2016
CY	4-17-C20B (CP)	3	2017
EN	4-5-C20B	20	2020
FA	2-6-C20B	6	2022
FI	7-14-C20B	7	2022
IN	2-7-C20B	6	2022
MI	3-30-C20B	9	2020
MP	7-19-C20B	18	2019
MS	6-8-C20B (MS70B/67J)	1	2017
OD	4-9-C20B	11	2017
QM	8-10-C20B	2	2014
SC	4-11-C20B (CP)	7	2022
TC	8-55-C20B	9	2018

For the following graph, the “% of Pure Instruction Hours spent on Possibly Relevant Content” column was calculated by dividing the number of possibly relevant pure instruction hours, by the sum of not relevant and possibly relevant pure instruction hours. For example, for AG: $66.4 + 800.1 = 866.5$, thus $66.4/866.5 = 7.66\%$.

POI Qualitative Analysis Branch Summary						
Branch	Lesson Relevance	Pure Instruction Hours	Total Classroom Hours (included for reference only)	Difference (included for reference only)	Number of Data Relevant Lesson Titles	% of Pure Instruction Hours spent on Possibly Relevant Content
ADA	Not Relevant	5145.6	5279.7	-134.1		
ADA	Possibly Relevant					
AG	Not Relevant	778.6	801.6	-23		
AG	Possibly Relevant	87.9	89	-1.10	5	10.14%
AR	Not Relevant	16,509.8	16,707.6	-197.8		
AR	Possibly Relevant					
AV	Not Relevant	608.5	630.3	-21.8		
AV	Possibly Relevant					
CBRN	Not Relevant	2,212.7	2,274.93	-62.2		
CBRN	Possibly Relevant	82.6	84	-1.4	2	3.6%
CY	Not Relevant	3,641.8	3,787.1	-145.3		
CY	Possibly Relevant	2,494.2	2,508.8	-14.6	18	40.65%
EN	Not Relevant	3,793.6	3,858	-64.4		

EN	Possibly Relevant					
FA	Not Relevant	13,982	13,982	-737.8		
FA	Possibly Relevant	243.2	256	-12.8	3	1.8%
FI	Not Relevant	936.5	980	-43.5		
FI	Possibly Relevant	4.6	6	-1.4	2	.49%
IN	Not Relevant	23,893.2	24,049.8	-156.6		
IN	Possibly Relevant					
MI	Not Relevant	4,510.8	4,649.6	-138.8		
MI	Possibly Relevant	68	73	-5	1	1.49%
MP	Not Relevant	2,869.6	2,928	-58.4		
MP	Possibly Relevant					
MS	Not Relevant	377.9	377.9	0		
MS	Possibly Relevant					
OD	Not Relevant	1,240.9	1,269.9	-29		
OD	Possibly Relevant	11.2	12	-0.8	2	.89%
QM	Not Relevant	1,366.3	1,396.76	-30.5		
QM	Possibly Relevant					
SC	Not Relevant	2,382.5	5,293.73	-2,911.23		

SC	Possibly Relevant	193.2	204.45	-11.25	10	7.5%
TC	Not Relevant	1,859.8	1,881.6	-21.8		
TC	Possibly Relevant					
<p>“Instruction hours” have been calculated by removing all “introduction” and “conclusion summary” contact hours, as well as any lesson title that contains “capstone” or “exercise”. This researcher believes this is a more accurate reflection of the content that students receive. For transparency, and to ensure that these numbers align with the numbers on each corresponding POI, both total hours and the difference have been articulated in this graph.</p>						

Lesson titles are Controlled Unclassified Information (CUI).

Full table is available at the following CAC enabled link:

https://www.milsuite.mil/wiki/A_Recipe_For_Data_Literacy:_Appendix_D_-_Data_Literacy_Relevant_Lesson_Titles

Lessons Identified as Relevant to Data Literacy (43 Total)				
Branch	Lesson Title	Instruction Hours	Total Hours	Difference
AG	Redacted	31.8	32.1	-0.3
AG	Redacted	8.4	8.8	-0.4
AG	Redacted	13.1	13.2	-0.1
AG	Redacted	29.4	29.5	-0.1
AG	Redacted	5.2	5.4	-0.2
CBRN	Redacted	79	80	-1
CBRN	Redacted	3.6	4	-0.4
CY	Redacted	17.6	18	-0.4
CY	Redacted	20.4	20.8	-0.4
CY	Redacted	20.8	21.2	-0.4
CY	Redacted	22	22.6	-0.6
CY	Redacted	7.2	7.6	-0.4
CY	Redacted	15.2	15.6	-0.4
CY	Redacted	158	160	-2
CY	Redacted	398.5	400	-1.5
CY	Redacted	398.5	400	-1.5
CY	Redacted	270.8	271.4	-0.6
CY	Redacted	109.2	109.6	-0.4
CY	Redacted	99.6	100	-0.4
CY	Redacted	271.2	271.6	-0.4
CY	Redacted	398.5	400	-1.5
CY	Redacted	197.5	200	-2.5
CY	Redacted	24	24.4	-0.4
CY	Redacted	30.8	31.2	-0.4
CY	Redacted	34.4	34.8	-0.4
FA	Redacted	60.8	64	-3.2
FA	Redacted	60.8	64	-3.2
FA	Redacted	121.6	128	-6.4
FI	Redacted	1.6	2	-0.4
FI	Redacted	3	4	-1
MI	Redacted	68	73	-5
OD	Redacted	5.6	6	-0.4

OD	Redacted	5.6	6	-0.4
SC	Redacted	76	80	-4
SC	Redacted	7.6	8	-0.4
SC	Redacted	12	16	-4
SC	Redacted	7.6	7.8	-0.2
SC	Redacted	5.2	5.6	-0.4
SC	Redacted	15.6	16	-0.4
SC	Redacted	6.8	7.2	-0.4
SC	Redacted	31.6	32	-0.4
SC	Redacted	15.2	15.9	-0.65
SC	Redacted	15.6	16	-0.4

APPENDIX E:

PERMISSION MEMOS

Permission to Review Basic Officer Leaders Course (BOLC)
Programs of Instruction (POIs)

ATZL-AUM

SUBJECT: Request to review the Basic Officer Leaders Course



DEPARTMENT OF THE ARMY
ARMY MANAGEMENT STAFF COLLEGE
250 GIBBON AVENUE, BLDG. 120
FORT LEAVENWORTH, KANSAS 66027-2314

ATZL-AUM

01 Jan 2022

MEMORANDUM FOR Matthew J. Bowman, LTC, AV, Chief, Basic Officer Leaders Course Division, HQ, U.S. Army Center for Initial Military Training, 210 Dillon Circle, Fort Eustis, VA 23604-5701

SUBJECT: Request acknowledgement of limited review of elements of the BOLC Basic (Phase 1) course information as part of research work of CPT Donald Ingham, a student at CGSC, enrolled in the Master of Military Arts and Science (MMAS) program. CPT Donald Ingham's MMAS research paper.

1. CPT Ingham is conducting his research on data visualization. Specifically, he is looking training time of data processing comparatively to other skills taught during initial training of Army Officers. CPT Ingham is triangulating this research by (1) surveying CGSC officers in order to understand their opinions of data processing training and abilities, (2) understanding the human cognitive processing of raw to outcome data, and (3) to using a review of the current state of foundational training in data processing experienced by new officer in the Army. The relevance is that Army doctrine and guidance includes an expectation, and as the Army strengthen modernization strategies, this will become a valuable and needed skill.
2. CPT Ingham is requesting to review the Basic Officer Leaders Course (BOLC) POIs and SharePoint documents (LTC Bowman recommended), in order to capture time allotted for specific topics and skills.
3. This memo will act as an assurance that no part of the actual curriculum will be reviewed or critiqued as part of the research; CPT Ingham is exclusively interested in time allows. As the presiding research Chair, I, Dr. Lynn M .Soots will be monitoring and ensuring CPT Ingham follows the parameters set out in this memo.
4. The point of contact is Dr. Lynn Soots, AMSC QAO Officer, AMSC Quality Assurance Office, (386)506-6133, lynn.m.soots.civ@army.mil.

ATZL-AUM

SUBJECT: Request to review the Basic Officer Leaders Course

5. Please acknowledge this memo below by digitally signing the memo.

 Digitally signed by
BOWMAN, MATTHEW JOSEPH 1
071841762
Date: 2022.01.05 15:35:57 -0800

MATHEW J. BOWMAN

LTC, AV

Chief, Basic Officer Leaders Course

Initial Internal Review Board (IRB) Survey Approval



REPLY TO
ATTENTION OF

DEPARTMENT OF THE ARMY
U.S. ARMY COMBINED ARMS CENTER
U.S. ARMY COMMAND AND GENERAL STAFF COLLEGE
100 STIMSON AVENUE
FORT LEAVENWORTH, KANSAS 66027-2301

ATZL-LSA-H

February 10, 2022

MEMORANDUM FOR: MAJ Donald F. Ingham, U.S. Army Command and General Staff College, Fort Leavenworth, KS 66027

SUBJECT: Approval to Conduct Human Subjects Research

1. Your protocol to research, **CA-2022-114** "*A Recipe for Data Literacy: Modernizing the Human Domain*" was finalized on 10 February 2022, was reviewed on 10 February 2022 and has been determined to be exempt from Institutional Review Board (IRB) review. As your proposal involves a survey of the of the CGSOC class of 2022, exploring "*Is there a deficiency between the number of hours Officers are trained on data processing and mid-level Officers perception of their own personal abilities to process data?*" you meet *Exemption #2.(i) of 32 CFR 219.104(d)*: "Research that only includes interactions involving educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior (including visual or auditory recording)" and that "the information is recorded by the investigator in such a manner that the identity of the human subjects cannot be readily ascertained, directly or through identifiers linked to the subjects."
2. Your protocol is approved with no required modifications. You have been assigned protocol approval number **22-114**. Reference this number when submitting any additional documentation or requesting information from the CGSC Human Protections Director concerning your research proposal.
3. Your proposal is approved for data collection with the following (5) stipulations:
 - (1) You are approved to recruit (744) student participants as indicated in your protocol. Any modifications to this study (including, but not limited to, changes in recruitment materials or procedures, investigators, inclusion/exclusion criteria, interview/survey questions, or data collection procedures, or increases in the number of participants enrolled) must be submitted as a written amendment for review and approval prior to implementing the change.
 - (2) You are expected to securely maintain all research documents and data collected for three (3) years.
 - (3) Allegations of non-compliance during this study will be investigated by the U.S. Army Command and General Staff College (CGSC), Human Protections Director (HPD).

ATZL-LSA-H

SUBJECT: Approval to Conduct Human Subjects Research (Ingham)

- (4) You are subject to auditing by the CGSC Human Protections Director (HPD) to ensure compliance. Failure to follow these guidelines could result in the termination of the approval for your research.
- (5) You are required to submit a Study Closure Report to the CGSC HPD upon completion of your study.

4. You must acknowledge the requirements of this approval letter, agreeing that the (5) stipulations are accepted by you. POC is the undersigned at michelle.a.miller48.civ@army.mil at (913) 684-7311 or (913) 682-1185.

MILLER.MICHELLE.ANN.11239310
48

Digitally signed by
MILLER.MICHELLE.ANN.1123931048
Date: 2022.02.10 11:00:46 -06'00'

MICHELLE A. MILLER, PhD
Human Protections Director
U.S. Army Command and General Staff College

Research Modification IRB Approval



REPLY TO
ATTENTION OF

DEPARTMENT OF THE ARMY
U.S. ARMY COMBINED ARMS CENTER
U.S. ARMY COMMAND AND GENERAL STAFF COLLEGE
100 STIMSON AVENUE
FORT LEAVENWORTH, KANSAS 66027-2301

ATZL-LSA-H

April 11, 2022

MEMORANDUM FOR: MAJ Donald F. Ingham, U.S. Army Command and General Staff College, Ft. Leavenworth, KS 66027

SUBJECT: Modification of CA-2022-114 Exempt Research Approval (Ingham)

1. Following careful review of your request for modification of CA-2022-114 "*A Recipe for Data Literacy: Modernizing the Human Domain*" dated 8 April 2022, I have determined the requested change does not alter its exempt status. The study remains exempt under 32 CFR 219.104(d)(2)(i.). The requested modification involves minor changes to the 1) problem statement, 2) updated thesis and bibliography, 3) new research questions, 4) new hypothesis for research question 1 one and 5) an application of new statistical tests.

2. The modification has been approved for the (5) changes listed in paragraph 1 above. This modification is in effect on 11 APR 2022. The study has been assessed as "No Greater Than Minimal Risk".

3. You are expected to comply with all conditions indicated in this memorandum and your previous approval memorandum, dated 10 FEB 2022, and to follow your approved protocol. You are subject to auditing by the CGSC Human Protections Director (HPD) to ensure compliance. Failure to follow these guidelines could result in the termination of the approval for your research. Any further modifications to this study (including, but not limited to changes in recruitment materials or procedures, investigators, inclusion/exclusion criteria, interview/survey questions, data collection procedures, or increases in the number of participants enrolled) must be submitted as a written modification for review and approval prior to implementing the change. Allegations of non-compliance will be investigated by the CGSC HPD.

4. You are responsible for the conduct of this study and for fulfilling the reporting requirements of 32 CFR 219, DoDI 3216.02, and submitting a Study Closure Report to the CGSC HPD upon completion of your study using Cayuse. POC is the undersigned at michelle.a.miller48.civ@army.mil at (913) 684-7311 or (913) 682-1185.

MILLER.MICHELLE.ANN.1123931048 Digitally signed by MILLER.MICHELLE.ANN.1123931048
Date: 2022.04.11 11:03:18 -05'00'

MICHELLE A. MILLER, PhD
Human Protections Director
U.S. Army Command and General Staff College

APPENDIX F:

DA PAM 600-3 SMARTBOOK DOCUMENT ANALYSIS

Branch	Branch/Functional Area Acronym	Branch Category	Knowledge and Skills	Word to search	# of times Data is mentioned
Air Defense Artillery	ADA	Operations	Redacted	data	0
Armor	AR	Operations	Redacted	data	0
Aviation	AV	Operations	Redacted	data	1
Signal Corps	SC	Operations Support	Redacted	data	0
Finance Corps	FI	Force Sustainment	Redacted	data	4
Engineer	EN	Operations	Redacted	data	0
Chemical	CBRN	Operations	Redacted	data	0
Civil Affairs	CA	Operations	Redacted	data	0
Field Artillery	FA	Operations	Redacted	data	0
Infantry	IN	Operations	Redacted	data	0
Military Police	MP	Operations	Redacted	data	0
Military Intelligence	MI	Operations Support	Redacted	data	0
Operations Research	ORSA	Operations Support	Redacted	data	8
Adjutant General	AG	Force Sustainment	Redacted	data	0
Logistics	LG	Force Sustainment	Redacted	data	0
Ordinance	OD	Force Sustainment	Redacted	data	0
Quartermaster	QM	Force Sustainment	Redacted	data	0
Transportation	TC	Force Sustainment	Redacted	data	0
Cyber	CY	Information Dominance	Redacted	data	0
Army Medical Department	MS	Special Branches	Redacted	data	0

APPENDIX G:

POWER BI MEASURES

Table	Measure Name	Calculation
BOLC Data Table	% of Each Class	[Total Instruction Hours]/calculate('BOLC Data Table'[Total Instruction Hours], ALLEXCEPT('What LookUp Table','What LookUp Table'[Chart Color Qual]))
BOLC Data Table	% of Hours Spent on Data	if(ISBLANK([Relevant Quant Hours]/[Total Instruction Hours]),"0%",[Relevant Quant Hours]/[Total Instruction Hours])
BOLC Data Table	% of Hours Spent on Data Qual	if(ISBLANK([Relevant Qual Hours]/[Total Instruction Hours]),"0%",[Relevant Qual Hours]/[Total Instruction Hours])
BOLC Data Table	Total Course Hours	Sum('BOLC Data Table'[ICH])
BOLC Data Table	Total Difference	[Total Instruction Hours]-[Total Course Hours]
BOLC Data Table	Total Instruction Hours	calculate([Total Course Hours],(FILTER('BOLC Data Table','BOLC Data Table'[Combined ELO/LSO]<>"Intro/Summary"&&'BOLC Data Table'[Lesson Title]<>"CAPSTONE")))
Smartbook Data Table	Sum of Data Mentions	sum('Smartbook Data Table'[# of times Data is mentioned])
Survey Data Table	Total Enough DL	Count('Survey Data Table'[Enough Data Literacy Education/Training in PME])
Survey Data Table	% Desired	[Not Enough DL]/'Survey Data Table'[Total Enough DL]
What LookUp Table	Count Data	if(isblank(Count('What LookUp Table'[Data In Lesson Title])),0,COUNT('What LookUp Table'[Data In Lesson Title]))
What LookUp Table	Count Information	if(isblank(Count('What LookUp Table'[Information In Lesson Title])),0,COUNT('What LookUp Table'[Information In Lesson Title]))
What LookUp Table	Count Lessons	DISTINCTCOUNT('What LookUp Table'[Lesson Title])
What LookUp Table	Count Process	if(isblank(Count('What LookUp Table'[Process In Lesson Title])),0,COUNT('What LookUp Table'[Process in Lesson title]))
What LookUp Table	Count Qual #	COUNT('What LookUp Table'[Chart Color Qual #])
What LookUp Table	Relevant Qual Hours	Calculate([Total Instruction Hours],(FILTER('What LookUp Table','What LookUp Table'[Chart Color Qual]="Possibly Relevant")))
What LookUp Table	Relevant Quant Hours	Calculate([Total Instruction Hours],(FILTER('What LookUp Table','What LookUp Table'[Chart Color Quant]="Possibly Relevant")))

APPENDIX H:

ARMY BRANCH ACRONYMS

ADA	Air Defense
AG	Adjutant General
AR	Armor
AV	Aviation
CA	Civil Affairs
CBRN	Chemical
CH	Chaplain
CY	Cyber
EN	Engineer
EOD	EOD
FA	Field Artillery
FA40	Army Space Operations Officer
FA46	Public Affairs
FA58	Enterprise Marketing and Behavioral Economics
FA59	Army Strategist
FI	Finance
IN	Infantry
IO	Information Operations
JAG	Judge Advocate General
MI	Military Intelligence
MP	Military Police
MS	Medical
OD	Ordinance Corps
ORSA	Operations Research/Systems Analysis
PSYOP	Psychological Operations
QM	Quartermaster Corps
SC	Signal
SOF	Special Forces
TC	Transportation
LG	Logistics

BIBLIOGRAPHY

- Association of College & Research Libraries. "Framework for Information Literacy for Higher Education (ACRL)." 2015. <http://www.ala.org/acrl/standards/ilframework>.
- Black, Ken. *Business Statistics for Contemporary Decision Making*. 6th ed. Hoboken, NJ: Wiley, 2009.
- Bronk, Justin. "The Mysterious Case of the Missing Russian Air Force." RUSI, February 28, 2022. <https://rusi.org/explore-our-research/publications/commentary/mysterious-case-missing-russian-air-force>.
- CGSC Learning Resource Center. Combined Arms Research Library. E-mail submission. February 27, 2022. Reviewed for grammar, punctuation, and clarity of expression.
- Creswell, John, and Vicki Plano. *Designing and Conducting Mixed Methods Research*. Thousand Oaks: SAGE Publications, 2017.
- Duffin, Erin. "Total military personnel of the U.S. Army for fiscal years 2020 to 2022, by rank." Statista, June 21, 2021. <https://www.statista.com/statistics/239383/total-military-personnel-of-the-us-army-by-grade/>.
- Dai, Yun. "How Many Ways Can We Teach Data Literacy?" *IASSIST Quarterly* 43, no. 40 (January 2, 2020): 1–11.
- Duignan, Brian. "Dunning-Kruger effect." Encyclopedia Britannica, September 8, 2020. <https://www.britannica.com/science/Dunning-Kruger-effect>.
- Few, Stephen. "Data Visualization Past, Present, and Future." Perceptual Edge, January 10, 2007. https://www.perceptualedge.com/articles/Whitepapers/Data_Visualization.pdf.
- Headquarters, Department of the Army. Army Doctrine Publication 5-0, *The Operations Process*. Washington, DC: Army Publishing Directorate, 2019.
- . Army Doctrine Publication 6-0, *Mission Command*. Washington, DC: Army Publishing Directorate, 2019.
- . *Army Multi-Domain Transformation: Ready to Win in Competition and Conflict*. Chief of Staff Paper #1. Unclassified version. March 16, 2021. <https://api.army.mil/e2/c/downloads/2021/03/23/eeac3d01/20210319-csa-paper-1-signed-print-version.pdf>.
- . Department of the Army Pamphlet 600-3, *Officer Professional Development and Career Management*. Washington, DC: Army Publishing Directorate, 2019.

- . Field Manual 6-0, *Commander and Staff Organization and Operations*. Washington, DC: Army Publishing Directorate, 2014.
- . Field Manual 6-02, *Signal Support to Operations*. Washington DC: Army Publishing Directorate, 2019.
- Kirk, Andy. *Data Visualisation*. Los Angeles: Sage, 2019.
- Morrow, Jordan. *Be Data Literate*. London: Kogan Page, Limited, 2021.
- Nussbaumer Knaflic, Cole, and Catherine Madden. *Storytelling with Data*. Hoboken, NJ: Wiley, 2020.
- Office of the Secretary, Department of Health, Education, and Welfare. *The Belmont Report: Ethical principles and Guidelines for the Protection of Human Subjects of Research*. Washington, DC: U.S. Department of Health and Human Services, April 18, 1979. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html>.
- Ogiela, Lidia, and Marek R Ogiela. *Cognitive Techniques in Visual Data Interpretation*. Studies in Computational Intelligence 228. Berlin, Heidelberg: Springer, 2009.
- Qlik. “Lead with Data: How to Drive Data Literacy in the Enterprise.” 2018. https://www.qlik.com/us/bi/-/media/08F37D711A58406E83BA8418EB1D58C9.ashx?ga-link=datlitreport_resource-library.
- QuestionPro. “Convenience Sampling: Definition, Advantages and Examples.” (Blog). Accessed March 29, 2022. <https://www.questionpro.com/blog/convenience-sampling/>.
- Usova, Tatiana, and Robert Laws. “Teaching a One-Credit Course on Data Literacy and Data Visualisation” *Journal of Information Literacy* 15, no. 1 (January 2021): 84–95. <https://www.ariadne.ac.uk/JIL/article/view/PRJ-V15-I1-3>.
- Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. *The Big Book of Dashboards*. New York: John Wiley & Sons, Incorporated, 2017.