



**NAVAL  
POSTGRADUATE  
SCHOOL**

**MONTEREY, CALIFORNIA**

**THESIS**

**COMPARISON OF ARTIFICIAL INTELLIGENCE  
METHODS TO ENHANCE AN AUTOMATED  
PEER-EVALUATION SUITE**

by

Andrew E. Nelson

September 2020

Thesis Advisor:  
Co-Advisor:

Neil C. Rowe  
Arijit Das

**Approved for public release. Distribution is unlimited.**

**THIS PAGE INTENTIONALLY LEFT BLANK**

<b>REPORT DOCUMENTATION PAGE</b>			<i>Form Approved 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 instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the 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 Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.				
<b>1. AGENCY USE ONLY (Leave blank)</b>		<b>2. REPORT DATE</b> September 2020		<b>3. REPORT TYPE AND DATES COVERED</b> Master's thesis
<b>4. TITLE AND SUBTITLE</b> COMPARISON OF ARTIFICIAL INTELLIGENCE METHODS TO ENHANCE AN AUTOMATED PEER-EVALUATION SUITE				<b>5. FUNDING NUMBERS</b>
<b>6. AUTHOR(S)</b> Andrew E. Nelson				
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> Naval Postgraduate School Monterey, CA 93943-5000			<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> N/A			<b>10. SPONSORING / MONITORING AGENCY REPORT NUMBER</b>	
<b>11. SUPPLEMENTARY NOTES</b> The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.				
<b>12a. DISTRIBUTION / AVAILABILITY STATEMENT</b> Approved for public release. Distribution is unlimited.			<b>12b. DISTRIBUTION CODE</b> A	
<b>13. ABSTRACT (maximum 200 words)</b>  A Department of Defense strategic focus area for artificial intelligence is the better allocation of personnel resources. The current peer-evaluation system at the Marine Officer Candidates School could benefit from artificial intelligence methods to partially automate the process. The school identifies performance trends by summarizing peer inputs and providing useful feedback to candidates to improve performance. This thesis used data from a recent training company and applied natural-language processing to preprocess peer inputs, identified phrases most helpful in predicting overall performance, extracted the best sentences for characterizing a candidate, and assembled draft counseling documents that required minimal revision by staff. Experiments with a prototype of our methods on a sample of real peer evaluations and summary counseling documents showed good though not perfect performance.				
<b>14. SUBJECT TERMS</b> peer evaluation, performance feedback, United States Marine Corps, USMC, Officer Candidates School, OCS, counseling, artificial intelligence, database, data synthesis, entry level training.			<b>15. NUMBER OF PAGES</b> 99	
			<b>16. PRICE CODE</b>	
<b>17. SECURITY CLASSIFICATION OF REPORT</b> Unclassified	<b>18. SECURITY CLASSIFICATION OF THIS PAGE</b> Unclassified	<b>19. SECURITY CLASSIFICATION OF ABSTRACT</b> Unclassified	<b>20. LIMITATION OF ABSTRACT</b> UU	

THIS PAGE INTENTIONALLY LEFT BLANK

**Approved for public release. Distribution is unlimited.**

**COMPARISON OF ARTIFICIAL INTELLIGENCE METHODS TO  
ENHANCE AN AUTOMATED PEER-EVALUATION SUITE**

Andrew E. Nelson  
Major, United States Marine Corps  
BS, U.S. Naval Academy, 2007

Submitted in partial fulfillment of the  
requirements for the degree of

**MASTER OF SCIENCE IN COMPUTER SCIENCE**

from the

**NAVAL POSTGRADUATE SCHOOL  
September 2020**

Approved by: Neil C. Rowe  
Advisor

Arijit Das  
Co-Advisor

Gurminder Singh  
Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

## **ABSTRACT**

A Department of Defense strategic focus area for artificial intelligence is the better allocation of personnel resources. The current peer-evaluation system at the Marine Officer Candidates School could benefit from artificial intelligence methods to partially automate the process. The school identifies performance trends by summarizing peer inputs and providing useful feedback to candidates to improve performance. This thesis used data from a recent training company and applied natural-language processing to preprocess peer inputs, identified phrases most helpful in predicting overall performance, extracted the best sentences for characterizing a candidate, and assembled draft counseling documents that required minimal revision by staff. Experiments with a prototype of our methods on a sample of real peer evaluations and summary counseling documents showed good though not perfect performance.

THIS PAGE INTENTIONALLY LEFT BLANK

# TABLE OF CONTENTS

<b>I.</b>	<b>INTRODUCTION.....</b>	<b>1</b>
<b>II.</b>	<b>RELATED WORK.....</b>	<b>5</b>
	<b>A. MULTI-SOURCE ASSESSMENT AND FEEDBACK.....</b>	<b>5</b>
	<b>B. DATA STATEMENTS.....</b>	<b>5</b>
	<b>C. SHORT-TEXT TRAINING.....</b>	<b>6</b>
<b>III.</b>	<b>PROBLEM SCOPE.....</b>	<b>9</b>
	<b>A. CURRENT OCS PEER EVAL SYSTEM.....</b>	<b>9</b>
	<b>B. SCOPED PROBLEM DEFINITION.....</b>	<b>14</b>
	<b>C. OTHER RELATED WORK.....</b>	<b>15</b>
	<b>1. Feature Choice.....</b>	<b>15</b>
	<b>2. Paraphrase and Summarization Techniques.....</b>	<b>17</b>
	<b>3. Bayesian Methods.....</b>	<b>19</b>
	<b>D. SUMMARY.....</b>	<b>21</b>
<b>IV.</b>	<b>METHODOLOGY.....</b>	<b>23</b>
	<b>A. DATA COLLECTION.....</b>	<b>23</b>
	<b>B. DATA PREPARATION.....</b>	<b>25</b>
	<b>C. IDENTIFYING PREDICTORS.....</b>	<b>27</b>
	<b>1. Correlation.....</b>	<b>28</b>
	<b>2. Least Squares Regression.....</b>	<b>33</b>
	<b>D. GENERATING THE FINAL COUNSELING DOCUMENT.....</b>	<b>34</b>
	<b>E. NATURAL-LANGUAGE PROCESSING: TRAIT-SENTIMENT EXTRACTION.....</b>	<b>36</b>
	<b>F. NATURAL-LANGUAGE PROCESSING: SUMMARY-STATEMENT EXTRACTION.....</b>	<b>37</b>
	<b>G. COMPILING A FINISHED DOCUMENT.....</b>	<b>39</b>
<b>V.</b>	<b>RESULTS.....</b>	<b>43</b>
	<b>A. TRAIT-SENTIMENT EXTRACTION.....</b>	<b>43</b>
	<b>B. SUMMARY-STATEMENT EXTRACTION.....</b>	<b>45</b>
	<b>C. SUMMARY-STATEMENT BIN UPDATES.....</b>	<b>49</b>
	<b>D. SUMMARY.....</b>	<b>54</b>
<b>VI.</b>	<b>CONCLUSION.....</b>	<b>55</b>

<b>APPENDIX A. OCS DATA STATEMENT .....</b>	<b>57</b>
<b>APPENDIX B. DATABASE SCHEMA .....</b>	<b>61</b>
<b>APPENDIX C. COMPARISON OF FINAL COUNSELING STATEMENTS .....</b>	<b>65</b>
<b>APPENDIX D. THREE ABOVE, THREE WITHIN, AND THREE BELOW- STANDARDS SYSTEM-GENERATED COUNSELING OUTPUTS .....</b>	<b>67</b>
<b>LIST OF REFERENCES .....</b>	<b>77</b>
<b>INITIAL DISTRIBUTION LIST .....</b>	<b>81</b>

## LIST OF FIGURES

Figure 1.	Two Basic Approaches to Artificial Intelligence. Source: Allen (2020).....	3
Figure 2.	Short-Form OCS Peer-Evaluation Dataset Data Statement.....	6
Figure 3.	Candidate Peer-Evaluation Input Sheet. Source: OCS (2020).....	10
Figure 4.	Visualization of Current OCS Peer-Evaluation Workflow.....	11
Figure 5.	OCS Peer-Evaluation Candidate Interview Form Template. Adapted from OCS (2020). .....	12
Figure 6.	OCS General Performance Review. Source: OCS (2014).....	13
Figure 7.	Visualization of Legacy OCS Peer-evaluation Workflow.....	14
Figure 8.	Normalized Term Frequency Calculation. Source: Raschka (2017). .....	16
Figure 9.	Inverse Document Frequency Formula. Source: Raschka (2017). .....	17
Figure 10.	Paraphrase Example.....	18
Figure 11.	Summarization Example.....	19
Figure 12.	Class Conditional Probability Equation. Source: Raschka (2017). .....	20
Figure 13.	Thesis Extract, Transform and Load Workflow Diagram .....	24
Figure 14.	Comparison of Spellchecking Packages Running in a Jupyter Notebook.....	26
Figure 15.	NLTK Stop Words with Words in Boldface to be Removed from List .....	27
Figure 16.	Example of Goal Output Counseling Document .....	35
Figure 17.	Extraction Method for Trait-Sentiment Statements.....	36
Figure 18.	Thesis Peer-Evaluation System Overview.....	40
Figure 19.	Analytics, Extraction, and Compilation Pipeline.....	40

Figure 20. Database E-R Diagram .....61

Figure 21. Output Document from OCS Dataset .....65

Figure 22. System Generated Output Document.....66

## LIST OF TABLES

Table 1.	Predictor and Response Variables from Training Data for Correlation by Normalized Final Rank.....	29
Table 2.	Correlation Values for Sample N-Grams from Table 1 .....	30
Table 3.	Top Five Unigrams, Bigrams, and Trigrams Indicating Higher Ranked or Higher Performing Candidates Based on Pearson Correlation with Normalized Rank (Left) And Assessment (Right) .....	32
Table 4.	Top Five Unigrams, Bigrams, and Trigrams Indicating Lower Ranked or Lower Performing Candidates Based on Pearson Correlation with Normalized Rank (Left) And Assessment (Right) .....	33
Table 5.	Trait-Sentiment Extraction Example without Weighting Statement Ratings .....	43
Table 6.	Trait-Sentiment Extraction Example Weighting Statement Ratings by Number of N-Grams in Each Statement .....	44
Table 7.	Trait-Sentiment Extraction Example Weighting Statements by Average Number of Words for All Ranked Statements .....	45
Table 8.	Comparison of the OCS Provided and System-Generated Summary Statement Paragraphs.....	46
Table 9.	Comparison of the Effects of Each Model on Summary-Statement Extraction for One Candidate .....	48
Table 10.	Trait-Sentiment Prior Probabilities in Our Training Set for the 32 Classes.....	49
Table 11.	Class Conditional Probabilities Generated by Naïve Bayes Classifiers for Three Example Phrases, with Highest Values Highlighted .....	50
Table 12.	Issue with Manual Summary-Sentence Labeling: Summary Sentence was Never Associated with a Top-Six Trait-Sentiment Pair .....	51
Table 13.	Issue with Manual Summary-Sentence Labeling: Summary-Sentence Content is Ambiguous.....	52

Table 14. Trait-Sentiment Prior Probabilities for Each of Seven Locally Labeled Classes on OCS Provided Summary Sentences .....53

Table 15. Outliers in 7-Classes Test on the OCS-Provided Summary Sentences .....53

Table 16. Examples of Issues in 7-Classes Test on Locally Created Summary Sentences.....54

## LIST OF ACRONYMS AND ABBREVIATIONS

AI	artificial intelligence
A-PES	Automated Performance Evaluation System
CIF	Candidate Interview Form
DOD	Department of Defense
GPR	General Performance Review
HQMC	Headquarters United States Marine Corps
IDF	inverse document frequency
JAIC	Joint Artificial-Intelligence Center
JJDIDTIEBUCKLE	justice, judgment, decisiveness, initiative, dependability, tact, integrity, endurance, bearing, unselfishness, courage, knowledge, loyalty, enthusiasm
MCO	Marine Corps Order
MCRD	Marine Corps Recruit Depot
MRO	Marine Reported On
MSAF	multisource assessment and feedback
NLP	natural-language processing
NLTK	Natural Language Tool Kit
NROTC	Naval Reserve Officer Training Corps
OCS	Officer Candidates School
PES	Performance Evaluation System
TBS	The Basic School
TF	term frequency

THIS PAGE INTENTIONALLY LEFT BLANK

## **ACKNOWLEDGMENTS**

To my family, for supporting me both in person and on the phone, across country, and during late nights at the dinner table, I am truly blessed to have this level of love and support.

To Professor Das for doing his very best to educate me on the art of database design. While our database work did not go as far as we wanted, it was a tremendous learning experience for me and, more importantly, our late afternoon discussions were something I looked forward to weekly.

To Dr. Rowe, who provided a necessary amount of grounding and superb editorial professionalism, thank you for your candid feedback, detailed reviews, and steady support.

THIS PAGE INTENTIONALLY LEFT BLANK

## I. INTRODUCTION

In his 2019 Commandant's Planning Guidance, General Berger addressed deficiencies in the current Marine Corps talent management and its Performance Evaluation System (PES). He said that better talent management would "encourage those you need and want to stay" and a better PES would "provide the Marine Reported On (MRO) with an opportunity at self-improvement" (Berger, 2019, pp. 7–8). An MRO is a specific class of Marine who rates a periodic fitness report, and the PES governs the personnel who are evaluated, who evaluate, and who manage the execution of the system. However, the PES is not the only system for performance evaluation. The Marine Leader Development order in 2017 provided a framework for the conduct of teaching, coaching, counseling, and mentoring throughout the Marine Corps (Headquarters United States Marine Corps [HQMC], 2017). It directed the conduct of these four activities to develop Marines starting with entry-level training and throughout their careers.

Despite several orders detailing the importance of counseling and evaluation, and prescribing methods by which Marine Corps leaders can execute the various evaluation systems in place (HQMC, 1986; 2000; 2017; 2018), most Marines will agree that person-to-person leadership development occurs infrequently. Generally, without a forcing function, such as requirements for data input into a computer system, Marine leaders will not conduct individually tailored professional development. Currently, the only two Marine Corps counseling and evaluation systems that are digitally automated to some degree are proficiency and conduct reports for junior enlisted Marines through the Marine Online portal and fitness reports for Marine officers through the Automated Performance Evaluation System (A-PES). Periodic counseling of subordinates is done on paper, if at all, and no digital system automates and archives these reports. Entry-level peer-evaluation systems are executed digitally to varying degrees but no standards for their development nor implementation exist. Fully digitizing and automating counseling and peer-evaluation could significantly improve the efficiency of the underlying systems, but more importantly, could help ensure the Commandant's intent is met to better develop Marine Corps talent.

Combining peer evaluations with performance counseling and evaluation is hotly debated. Using peer-evaluation data as part of a counseling document has been considered by the Army (Department of the Army, 2018; McAninch, 2016; Wike, n.d.), the Marine Corps (Nelson, 2019; Niedziocha, 2014), the Navy (Navy Personnel Command, 2019; Personnel Command Public Affairs, 2018), and civilian groups (Hardison et al., 2015; Lepsinger & Lucia, 1997; McCauley et al., 1997; Tirona & Gislason, 2011). Highlighted benefits include gaining insight from subordinates on a leader's performance (Niedziocha, 2014), reviewing a broader range of information to minimize bias (Lepsinger & Lucia, 1997), and building better developmental relationships between leaders and subordinates (McAninch, 2016). Possible disadvantages are qualitative peer remarks in the control of senior raters (Nelson, 2019), using peer-evaluation feedback for personnel appraisal instead of development (Lepsinger & Lucia, 1997; Tirona & Gislason, 2011), and potential breach of peer-evaluation anonymity (McCauley et al., 1997, p. 9). It has been argued that peer evaluations at Marine Corps Officer Candidates School (OCS) should not be directly used for performance evaluations, though the unique mission of OCS for the evaluation and screening of candidates provides an acceptable context.

Accepting that performance counseling using peer-evaluation data is desirable for the overall evaluation and screening process at OCS, we should try to modernize and streamline this process to take full advantage of digital tools. Upgrading to a digital system permits analytics to run on peer-evaluation data to generate counseling documents and enhance the situational awareness of the command.

One possibility is to use artificial intelligence in such a digital system. Artificial intelligence is a broad term that according to the U.S. Department of Defense “refers to the ability of machines to perform tasks that normally require human intelligence—for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action,” (Department of Defense [DOD], 2019). Several artificial intelligence concepts could apply to peer-evaluation systems such as the one used at OCS to streamline tasks and increase efficiency of the unit (DOD, 2019, p. 11). These methods fall under both approaches to artificial intelligence, as shown in Figure 1, though most methods analyzed in this thesis will fall under the machine-learning approach.

**Figure 1: Simplified Diagram of AI Approaches**

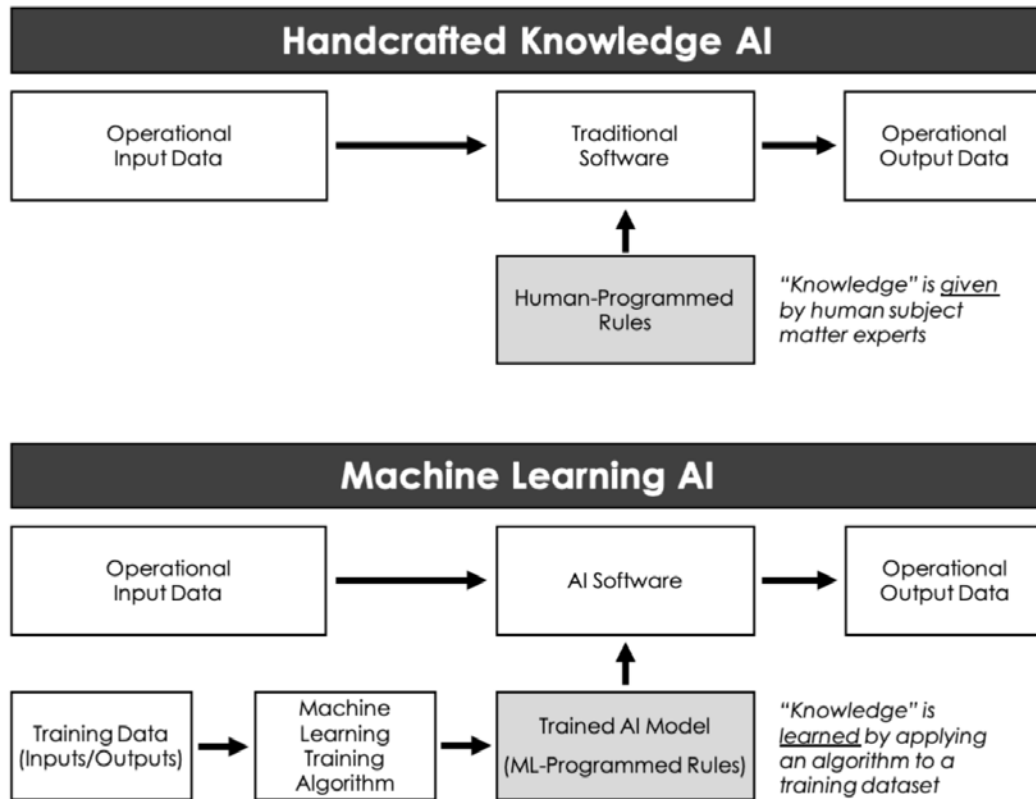


Figure 1. Two Basic Approaches to Artificial Intelligence.  
Source: Allen (2020).

Peer-evaluation data is primarily text with minimal structure. Natural-language processing is a branch of artificial intelligence that uses both rules-based and machine-learning techniques to interpret human language text and prepare it for further processing (SAS Institute Inc., 2020). Natural-language processing techniques could analyze raw peer-evaluation data to automate the extraction of information. Machine-learning could improve the identification of important and unique information over what a human evaluator could find manually. Classifying peer-evaluation inputs could aid us in creating useful counseling statements that summarize general peer issues and make it easier for a reviewer to finish the report, thereby reducing human processing time. Lastly, regression analysis could help us identify trends within the units conducting peer-evaluation sessions that may enhance review by command leadership.

It is valuable to the professional development of officer candidates to get personalized performance counseling based on their peer evaluations. Continuing to aggregate and process this information by hand is impractical and inefficient. Choosing the right artificial-intelligence techniques to process the peer-evaluation data and develop useful counseling reports for all candidates supports the Commandant's vision outlined in his planning guidance. Such techniques for a digital peer-evaluation system can be designed to engender trust in the system and adhere to the Department of Defense Artificial Intelligence Principles (Department of Defense Joint AI Center [JAIC], 2020). For the current OCS peer-evaluation process, the two tasks best augmented with artificial intelligence are picking the most useful statements to paraphrase aggregated inputs and selecting the best summary statements for counseling paragraphs. There is simply not enough time in the training cycle for a human staff member to review every input for each candidate, cross-reference, and compare them to find the most useful and distinctive information. Artificial-intelligence algorithms can accomplish this task in a fraction of the time. In addition, these algorithms can draft counseling statements which will help further speed the completion of quality documents for all candidates.

In the following chapters, we will review related work in the fields of performance counseling and artificial intelligence, refine the scope of the OCS peer-evaluation counseling problem, develop a methodology to use different artificial intelligence methods to augment the development of peer-evaluation counselings, discuss the results of the application of our methods on training data from a recent OCS company, and propose some potential future work in this area.

## **II. RELATED WORK**

### **A. MULTI-SOURCE ASSESSMENT AND FEEDBACK**

The importance of performance counseling has been discussed both within military circles (Criley, 2006; Edson, 1985; Garrett, 1996; Niedziocha, 2014) and civilian circles (Lepsinger & Lucia, 1997; McCauley et al., 1997) for decades alongside its counterpart, performance evaluation. Multisource Assessment and Feedback (MSAF) is a way of conducting either counseling or evaluation whereby inputs are asked from several people, reviewed, and compiled usually by one person, and then provided in some form to the person being counseled or evaluated. An example MSAF system is a “360-degree” feedback system that includes inputs from seniors, peers, and subordinates. Other MSAF systems include peer-evaluation systems used at Marine Corps entry-level officer training which take multiple peer inputs and include an overall assessment from the immediate supervisor. Much debate centers on how we should separate counseling from evaluation (Chadwick, 2013; Garrett, 1996; Lepsinger & Lucia, 1997; McCauley et al., 1997). Counseling could influence future work while evaluation covers past work. The Marine Corps directs that subordinates be counseled during the reporting period and evaluated at the end (HQMC 1986; 2018). Using the results of MSAF systems as direct inputs to evaluations is a practice frowned upon for legal and ethical reasons (Hardison et al., 2015; Lepsinger & Lucia, 1997). What is generally agreed on is that MSAF can be a useful tool for performance counseling, especially if it remains anonymous (Chadwick, 2013; Lepsinger & Lucia, 1997; McCauley et al., 1997).

### **B. DATA STATEMENTS**

To assure those using an artificial-intelligence system that uses natural-language processing that a good-faith effort has been made to reduce potential biases, presenting some explanatory information with each dataset is important. This technique can reduce the concern that the application of artificial-intelligence methods trained on data from one population do not apply to another population (Bender & Friedman, 2018). One proposal is to require data statements that will, “will bring about improvements in engineering and

scientific outcomes while also enabling more ethically responsive [natural-language processing] technology” (Bender & Friedman, 2018, p. 587). They argue that such data statements should be included in every natural-language processing system that uses a new dataset as well as any publication that reports the results of experiments (Bender & Friedman, 2018, p. 590). Our research used a loosely structured Marine Corps peer-evaluation dataset that we further refined and annotated. We propose additional information that will take the form of the short-form data statement in Figure 2, and a long-form data statement in Appendix A, following (Bender & Friedman, 2018, p. 591).

**Short Form Data Statement**

**OCS Peer-evaluation Dataset. This dataset includes ranking and assessment labels for 207 individual peer-evaluation sheets with labels for 1,242 peer-evaluation sheet trait inputs. It also includes ranking, assessment, trait labels for six traits, and a final counseling paragraph for seven final counseling output peer-evaluation forms. The trait labels indicate which of the 14 Marine Corps leadership traits were reflected by a candidate in either a positive or negative way according to one of their peers. This dataset represents most peer-evaluation inputs and outputs generated by an OCS training company for one peer-evaluation iteration during a Spring 2020 training cycle.**

Figure 2. Short-Form OCS Peer-Evaluation Dataset Data Statement

### C. SHORT-TEXT TRAINING

Traditional methods to process text rely on statistical methods that only work well with many words to analyze (Vaishnavi et al., 2013). For shorter texts such as sentences or tweets, insufficient statistics are available to compare two sentences and determine if they are similar enough to be paraphrases of each other. A new approach uses a semantic-similarity algorithm to identify grammar patterns in short text, which may then be compared based on their string similarity and a rating derived from WordNet (Princeton University, 2010; Vaishnavi et al., 2013). The authors propose a new similarity feature

space called “semantic coordinate space” in which to project the similarity values between grammar patterns of short texts. Summing the maximum similarity rating for each grammar pattern in the input with the fewest patterns yields an overall similarity rating which can be normalized for comparisons.

THIS PAGE INTENTIONALLY LEFT BLANK

### **III. PROBLEM SCOPE**

#### **A. CURRENT OCS PEER EVAL SYSTEM**

While formal orders governing the junior enlisted development (HQMC 2000), senior enlisted, and officer development (HQMC 2017; 2018) exist, no similar Marine Corps Order (MCO) governs officer-candidate and entry-level junior-officer performance counseling. Counseling and peer evaluations at Naval Reserve Officer Training Corps (NROTC) units, Officer Candidates School (OCS), and The Basic School (TBS), are executed following local standards using locally designed tools. No standard governs the systems that support these personnel, and often the units compile data by hand and do little analysis of it.

At OCS, peer evaluations are completed by hand on paper, and by all officer candidates, two to three times per training cycle. Each candidate writes one peer-evaluation sheet on every other candidate in their squad, as in Figure 3. The peer-evaluation sheet includes a squad ranking, a performance assessment (above, within, or below average), three traits picked from the fourteen Marine Corps Leadership Traits (Training Command, 2020) with short explanations about how the candidate has shown positive performance, and three traits with short explanations about how the candidate has shown negative performance. These fourteen traits (justice, judgment, decisiveness, initiative, dependability, tact, integrity, endurance, bearing, unselfishness, courage, knowledge, loyalty, and enthusiasm) expressed with explicit positive or negative sentiment provide candidates with 28 trait classes to choose from when rating their peers. We refer to the combination of a trait with its positive or negative sentiment as a “trait-sentiment pair.”

**SQUAD PEER EVALUATION I**

**1. Provide a rank of where you believe this candidate falls out in your squad.**

RANK		of	
------	--	----	--

**2. Provide at least (3) positive leadership traits you observed of this candidate during this evaluation period. Provide a detailed explanation or example of why you chose a specific trait.**


**3. Provide at least (3) negative leadership traits you observed of this candidate during this evaluation period. Provide a detailed explanation or example of why you chose a specific trait.**


**4. Circle the overall assessment of the candidate's performance during this period.**

ABOVE STANDARDS                     
  WITHIN STANDARDS                     
  BELOW STANDARDS

CANDIDATE EVALUATED (LAST NAME, FI. MI.)	COMPANY	PLATOON	DATE

Figure 3. Candidate Peer-Evaluation Input Sheet. Source: OCS (2020).

Training staff at the platoon level then review all submitted paper peer-evaluations, enter some evaluation data into an Excel spreadsheet to calculate rankings, and tally other data points by hand on a standard form to determine the overall performance assessment

and top three positive and three negative character traits for each candidate; Figure 4 shows the workflow. Next, a staff member reviews evaluations for each candidate again to extract relevant remarks about each top trait, and types them into a final counseling document. To complete the document, the staff member writes a statement summarizing the main things that the candidate is doing well and the main things they must work on to do better in the program.

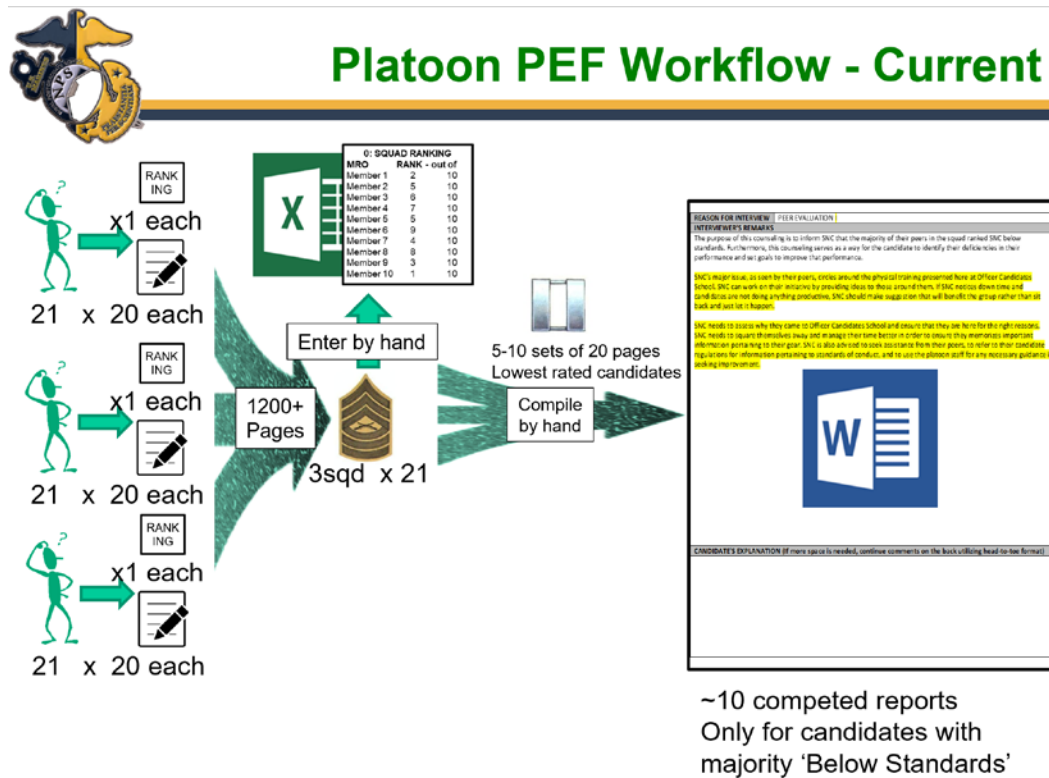


Figure 4. Visualization of Current OCS Peer-Evaluation Workflow

Presently, final counseling documents are only completed for those candidates judged by their peers to have performed “below standards”; the final document, called a Peer-evaluation Form, which is a special subtype of OCS standard Candidate Interview Form (CIF) (Figure 5) only includes limited information: an overall report performance assessment influenced by the suggested performance assessment marks from peers and a

final counseling statement that summarizes key areas in which the candidate can improve performance in the eyes of their peers with comments from their platoon staff supervisor.

CANDIDATE INTERVIEW FORM			
<b>REASON FOR INTERVIEW</b>	<b>PEER EVALUATION</b>		
<b>INTERVIEWER'S REMARKS</b>			
<p>The purpose of this counseling is to inform SNC that the majority of their peers in the squad ranked SNC below standards. Furthermore, this counseling serves as a way for the candidate to identify their deficiencies in their performance and set goals to improve that performance.</p> <p>SNC's major issue, as seen by their peers, circles around the physical training presented here at Officer Candidates School. SNC can work on their initiative by providing ideas to those around them. If SNC notices down time and candidates are not doing anything productive, SNC should make suggestion that will benefit the group rather than sit back and just let it happen.</p> <p>SNC needs to assess why they came to Officer Candidates School and ensure that they are here for the right reasons. SNC needs to square themselves away and manage their time better in order to ensure they memorizes important information pertaining to their gear. SNC is also advised to seek assistance from their peers, to refer to their candidate regulations for information pertaining to standards of conduct, and to use the platoon staff for any necessary guidance in seeking improvement.</p>			
<b>CANDIDATE'S EXPLANATION (If more space is needed, continue comments on the back utilizing head-to-toe format)</b>			
ABOVE STANDARDS	WITHIN STANDARDS	BELOW STANDARDS	
CANDIDATE'S SIGNATURE _____		CAPT LAST, FI. MI. _____	
<b>CANDIDATE INTERVIEWED (LAST NAME, FI. MI.)</b>	<b>COMPANY</b>	<b>PLATOON</b>	<b>DATE</b>
LASTNAME, FI. MI.	D	1	YYMMDD

Figure 5. OCS Peer-Evaluation Candidate Interview Form Template. Adapted from OCS (2020).

In the past, a final document called a General Performance Review (GPR) was created for all candidates, which included an overall ranking of the candidate as compared to their other squad members, an overall performance assessment influenced by the suggested performance assessment from each peer, the top three positive and three negative character traits as determined by their peers with a sample of peer remarks for each trait,

and a final counseling statement (Figure 6). The GPR added information for the candidate being counseled and could offer good benefits from system automation due to the larger number of personnel processing a significantly higher number of GPR documents. See Figure 7 for an example workflow of this legacy process.

**UNITED STATES MARINE CORPS**  
COMPANY 6  
OFFICER CANDIDATES SCHOOL  
TRAINING COMMAND  
2189 ELIROD AVENUE  
QUANTICO, VA 22134-5001

GENERAL PERFORMANCE REVIEW (GPR)

YYMMDD

**FAVORABLE                      MARGINAL                      UNFAVORABLE**

**REASON FOR INTERVIEW: PEER EVALUATION I**

**INTERVIEWERS REMARKS:** The purpose of this counseling is to inform the candidate that he was ranked **I** of **I** in his squad during peer evaluation **I**. SNC received **I** FAVORABLE, **I** MARGINAL, and **I** UNFAVORABLE evaluations. Furthermore, this counseling serves as a way for the candidate to identify both negative and positive trends in his performance and set goals to improve that performance.

**Negative Trends:**  
- **Trait.** Give **three** negative traits from the Peer Evals in this format.

**Positive Trends:**  
- **Trait.** Give **three** positive traits from the Peer Evals in this format.

**Goals to Improve/Sustain Performance:**  
//Insert comments that are appropriate to each candidate to help him/her improve performance. Do NOT copy and paste the same statement in each GPR without adding/editing to personalize.  
i.e. SNC Continues to demonstrate that he is uncomfortable in leadership positions. With more familiarity of the knowledge, SNC's confidence should improve. SNC cannot be afraid to make a decision. SNC needs to have stronger physical courage if he is to pass the hike and graduate from OCS. SNC is advised to review the above and strive for improvement in demonstrating continued positive traits and reducing the amount and severity of negative trends.

**CANDIDATE'S EXPLANATION (utilize back of CHIT for more room using head-toe format):**

\_\_\_\_\_  
Candidate's Signature

\_\_\_\_\_  
Capt I. M. TOUGHER

LAST **FI. MI.**                      **OCC-215**                      **C**                      **1st**  
SURNAME                      INITIALS                      CLASS                      CO                      PLT

Figure 6. OCS General Performance Review. Source: OCS (2014).



# Platoon GPR Workflow - 2014

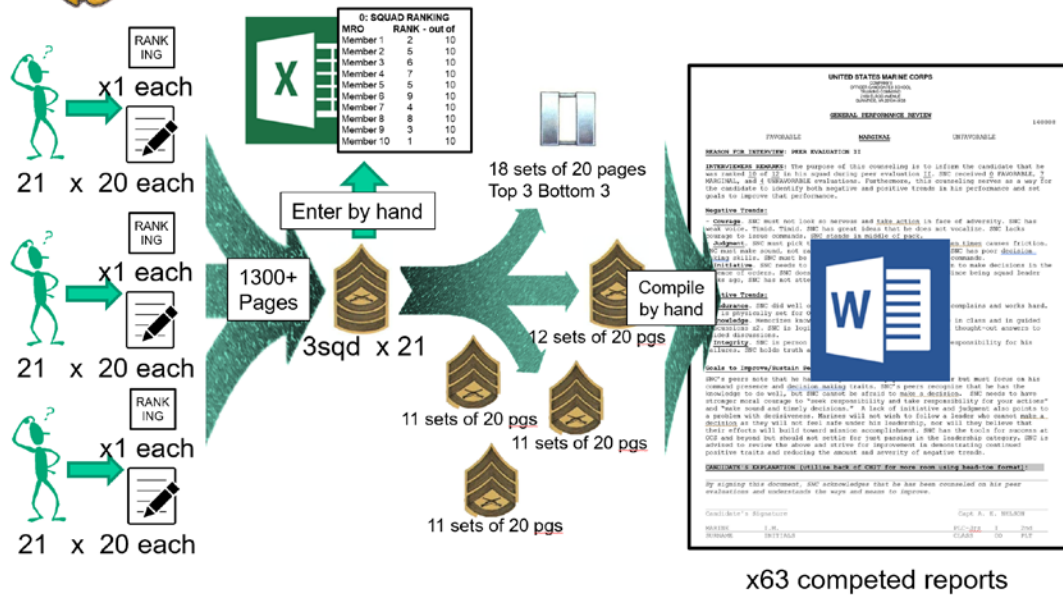


Figure 7. Visualization of Legacy OCS Peer-evaluation Workflow

## B. SCOPED PROBLEM DEFINITION

Previously, when OCS platoon staff counseled every candidate on peer-evaluation results using written GPRs, processing the roughly 1200 individual peer-evaluation sheets per platoon took more than 37 man-hours per iteration. The process was tedious, prone to be hurried, and the output lacked detail given the time constraints. The current peer-evaluation system only requires creating a Peer-evaluation Form and counseling low-ranked candidates. All other candidates receive only the raw peer-evaluation input sheets to review. The reduction in required man-hours for the current system comes at a cost of reducing counseling value to all candidates. Candid feedback requires anonymity (Lepsinger & Lucia, 1997; McCauley et al., 1997), and providing raw peer-evaluation documents directly to each candidate removes it from the process. Also, showing raw feedback directly to people who lack interpretive context will be unlikely to enable positive behavioral change (McAninch, 2016). Staff members have the training and experience to put key points in the proper context. A digital system that automates the compilation of

peer-evaluation input forms might achieve a similar reduction in man-hours to process documents while providing fair, anonymous counseling to every candidate.

Natural-language processing techniques could identify important concepts within a document or group of documents and potentially combine them to achieve some level of anonymity (Vaishnavi et al., 2013). Correlation and linear regression could identify text features that best predict the trait to which a given output statement best aligns, enabling the generation of final counseling statements. Pre-trained neural networks could derive semantic meaning to help compare sentences to paraphrase and summarize text (Freihat et al., 2013; Tiha, 2017).

## **C. OTHER RELATED WORK**

### **1. Feature Choice**

Features are central to data analysis (Nikhath & Subrahmanyam, 2019). Selecting the best features before analysis is critical to creating effective algorithms to classify and compare text documents.

The first methods applied to raw text input are called preprocessing. They include the removal of extraneous information, standardizing word case and punctuation, identification of unimportant words, and stemming or lemmatization (Tiha, 2017; Vaishnavi et al., 2013). Usually, text is first tokenized by breaking it down into sentences and then into single words or multi-word sequences (N-grams) consisting of N words in sequence (Tiha, 2017).

Words and word sequences that occur often do not make good features because they cannot distinguish sentences well. Therefore, the usual next step is to remove N-grams containing such common-use “stop” words and word sequences. The Natural Language Tool Kit (NLTK) (Bird et al., 2020) for the Python programming language includes a basic list of common stop words among many other basic natural-language processing capabilities. It also provides useful general tools for stemming and lemmatization. Stemming uses rules to remove the more common endings from words to get to the basic form of the word (Porter, 2001). This standardizes the form of related words and helps improve subsequent feature extraction. For example, the words “find,” “finding,” and

“findings” are all stemmed to the word “find.” Lemmatization does something similar but reduces words to a form found in a dictionary (Manning et al., 2009, pp. 32–34).

Once raw input text has been transformed through preprocessing, additional features may be derived through methods like frequency analysis, correlation, filtering, similarity analysis, or identifying keywords and cue phrases (Nikhath & Subrahmanyam, 2019; Tiha, 2017). Frequency analysis analyzes the number of times a word or N-gram appears in each document. Term frequency (TF) is the count of a word or N-gram in a document. Since documents vary in size from one to another, the normalized term frequency in Figure 8 is also used in this thesis when documents are larger and the variance between N-gram counts is larger.

$$\text{normalized term frequency} = \frac{tf(t, d)}{n_d}$$

where

- $tf(t, d)$ : Raw term frequency (the count of term  $t$  in document  $d$ ).
- $n_d$ : The total number of terms in document  $d$ .

Figure 8. Normalized Term Frequency Calculation.  
Source: Raschka (2017).

Inverse document frequency (IDF) is the number of documents in which a given word or N-gram appears within all documents in a collection given by the equation in Figure 9.

$$idf(t) = \log \left( \frac{n_d}{n_d(t)} \right),$$

where

- $n_d$ : The total number of documents.
- $n_d(t)$ : The number of documents that contain the term  $t$ .

Figure 9. Inverse Document Frequency Formula.  
Source: Raschka (2017).

Combining the two frequency values by multiplying them (called “TF-IDF”) is often done, and results in features that describe the importance and rarity of each term in one document from a set (Manning et al., 2009, p. 117; Tiha, 2017). Filtering methods can identify important features. Often features are ranked and a threshold value can be adjusted to limit the total number of features considered.

## 2. Paraphrase and Summarization Techniques

Paraphrasing rewrites text to make it easier to understand while preserving key points from the original text. The paraphrased content may be of similar length to the source content and it should follow the source content closely. (EduBirdie, 2019). In peer evaluation, some points will be redundant and should be condensed (Vaishnavi et al., 2013). Other points may distinguish a person from their peers and should be emphasized. The number of input sentences to paraphrase in a peer-evaluation system should increase with the number of peers and should be less than the length of a reasonable summary. Paraphrasing should reduce the text to a few sentences that capture the very best information available.

Word-based rating techniques can rate sentences based on the features extracted from them. The features most often used are term frequency, inverse document frequency, and TF-IDF for single words or multi-word N-grams (Tiha, 2017). The highest rating sentences can be selected as the best sentences for a paraphrase of a document. For an

example of paraphrasing, consider the three potential input sentences for the trait-sentiment pair “discipline-negative” in Figure 10. For term-frequency calculations we will choose not to normalize the raw term frequencies due to the short sentence lengths. For inverse-document calculations we will treat each sentence as a separate document.

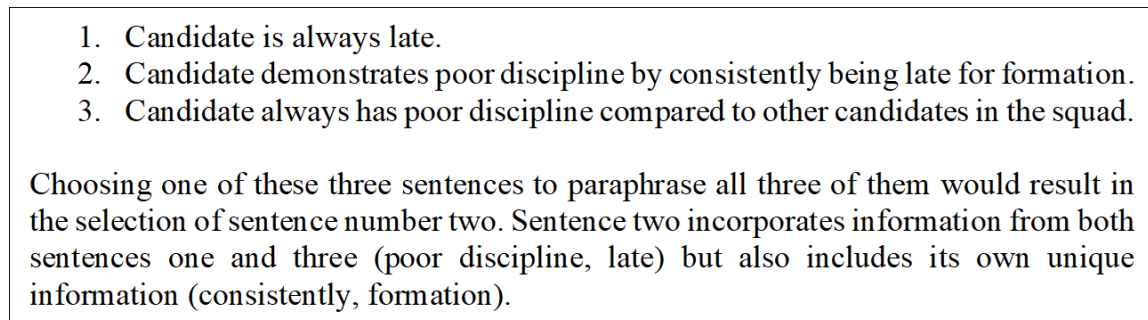


Figure 10. Paraphrase Example

Summarizing significantly reduces text compared to paraphrasing while keeping key points from the material. A summary is good when it omits “all the irrelevant information, including only those facts that matter to your discussion ... [and] you aim at briefly familiarizing readers with some content, without falling into discussion of irrelevant details” (EduBirdie, 2019). A good peer-evaluation summary reports key themes and patterns relevant to performance (Tirona & Gislason, 2011) coupled with an evaluator perspective on how to improve. Summarization can be extractive or abstractive. Extractive summarization identifies the important sentences (Tiha, 2017, p. 2) and combines them. Abstractive techniques eliminate redundancies in the text using grammar patterns to determine syntactic similarity (Vaishnavi et al., 2013), and then generate sentences that capture and clarify the meaning of the remaining information. Abstractive techniques are not as mature as extractive techniques (Tiha, 2017). For an example of extractive summarization, consider the three potential summary body sentences in Figure 11, again for a candidate with the trait-sentiment pair “discipline-negative.” In this case, we also know from other term-frequency analysis that the important N-grams are “formation,” “listen to,” and “failure to follow.”

1. Candidate can demonstrate better discipline by trying to listen to their billet holders instead of talking in formation.
2. Candidate is known for their poor bearing and failure to follow basic directions.
3. Candidate can increase discipline by practicing drill manual during free time.

Choosing one of these three sentences to be part of the summary would result in the selection of sentence number one. Sentence one directly corresponds to a known trait-sentiment pair and contains key information (discipline, listen to, formation). Sentence two contains information corresponding to a different trait-sentiment pair (bearing) and also contains one piece of key information (failure to follow), while sentence three contains just one piece of information that corresponds to the trait-sentiment pair (discipline) but contains no other key information.

Figure 11. Summarization Example

### 3. Bayesian Methods

Classifying sentences and features is aided by a calculation of the probability of combined relevance of all the N-grams. Bayes' theorem combined with the notion of conditional independence provides a simple but powerful method of classification called Naïve Bayes classification (Raschka, 2017). We chose to implement the multinomial method instead of the multi-variate Bernoulli drawing on the conclusions from (McCallum & Nigam, 1998). We used term-frequency analysis to derive probabilities for all classes we tried to discriminate and then found the class with the highest derived value according to formula (3.1).

$$\text{posterior probability} = \frac{\text{class conditional probability} * \text{class prior probability}}{\text{evidence}}$$

(3.1)

The posterior probably is the probability of a class given some evidence. The prior (initial) probability of a class can be estimated from the training data by counting samples of the class in the training set and dividing by the total number of samples. The evidence term is the same for each class, so it can be dropped as we are comparing classes (Raschka, 2017).

The class conditional probabilities are derived using the formula in Figure 12 and converted to log probabilities to handle potential numerical stability issues (Deshpande, 2017).

The term frequencies can then be used to compute the maximum-likelihood estimate based on the training data to estimate the class-conditional probabilities in the multinomial model:

$$\hat{P}(x_i | \omega_j) = \frac{\sum tf(x_i, d \in \omega_j) + \alpha}{\sum N_{d \in \omega_j} + \alpha \cdot V}$$

where

- $x_i$ : A word from the feature vector  $\mathbf{x}$  of a particular sample.
- $\sum tf(x_i, d \in \omega_j)$ : The sum of raw term frequencies of word  $x_i$  from all documents in the training sample that belong to class  $\omega_j$ .
- $\sum N_{d \in \omega_j}$ : The sum of all term frequencies in the training dataset for class  $\omega_j$ .
- $\alpha$ : An additive smoothing parameter ( $\alpha = 1$  for Laplace smoothing).
- $V$ : The size of the vocabulary (number of different words in the training set).

The class-conditional probability of encountering the text  $\mathbf{x}$  can be calculated as the product from the likelihoods of the individual words (under the *naive* assumption of conditional independence).

$$P(\mathbf{x} | \omega_j) = P(x_1 | \omega_j) \cdot P(x_2 | \omega_j) \cdot \dots \cdot P(x_n | \omega_j) = \prod_{i=1}^m P(x_i | \omega_j)$$

Figure 12. Class Conditional Probability Equation.  
Source: Raschka (2017).

Putting together our prior probabilities and class conditional probabilities, we derive formula (3.2) to calculate the posterior probably for a given class.

$$P(w_j | \mathbf{x}) = P(w_j) * P(\mathbf{x} | w_j) = P(w_j) * \prod_{i=1}^m P(x_i | w_j) \quad (3.2)$$

We can rewrite the equation as formula (3.3) with the logarithms of probabilities.

$$\log P(w_j | \mathbf{x}) = \log P(w_j) + \sum_{i=1}^m \log P(x_i | w_j) \quad (3.3)$$

Assuming conditional independence of features, the class conditional probability for a given set of features can be estimated as the sum of the log prior probability and the individual log conditional probabilities for features in the set given the class. Since some

features seen in a new object for classification may not exist in the training set, a smoothing factor or default count is usually included in the conditional probability calculation (Deshpande, 2017; Raschka, 2017). When comparing two or more predictions, the class with the highest value is the predicted class.

#### **D. SUMMARY**

This section reviewed the state of the OCS peer-evaluation system and examined where artificial-intelligence methods to augment automation could improve that process. Specifically, we reviewed choosing features that could be used during paraphrasing, summarization and in the implementation of a classifier. Next, we will discuss how we implemented these techniques in our work.

THIS PAGE INTENTIONALLY LEFT BLANK

## IV. METHODOLOGY

To study how to automate and improve the efficiency of the generation of effective counseling documents for candidates, our research analyzed peer evaluations and Peer-evaluation Form summaries from a recent OCS training company. It selected sentences describing the top positive and negative traits and ranked them for inclusion in counseling statements. We assumed:

- Sufficient training examples could cover each possible trait we want to extract from the data.
- In the output counseling paragraphs, only primary positive and negative traits were discussed except for opening and closing sentences.
- For ranking traits, a tie should be decided by the order in which the traits appear as initials in the Marine Corps acronym “JJIDITIEBUCKLE” (Training Command, 2020).

### A. DATA COLLECTION

Even when the OCS peer-evaluation process changed to the current shortened system, the type and format of input data remained the same. This data includes both quantitative and qualitative remarks that must be aggregated, tallied, filtered, processed, and finally put into context by platoon staff. One way to handle this type of data is a relational database. A relational database provides rules for a program or process to follow to access and manipulate the stored data, and provides data-integrity guarantees through a specification (Oracle Corporation, 2020a). For our system, neither the expected amount of data stored nor the number of simultaneous users required a big-data solution. Individual reports should only be worked on by one user at a time, and once complete, reports should be stored securely in a read-only manner. Input and output data should also be available for query by authorized parties such as the higher headquarters.

We implemented a simple prototype relational database using the Oracle Structured Query Language (SQL) Developer program, version 17.3.1.279 running on Java

1.8.0\_144. The database entity-relationship diagram and the Python code for loading from and updating the database are found in Appendix B. The database included tables for storing basic pre-populated data, system user data, peer-evaluation session data, candidate input data, and session output data. Data that could be pre-populated was loaded by a second set of SQL commands executed from within SQL Developer. Training and testing data provided from OCS were loaded into appropriate tables using SQL Developer's Actions-Import Data wizard. Data manipulation was handled with a Jupyter Notebook running Python 3.8 and connecting to the database by the "cx\_Oracle" Python package (Oracle Corporation, 2020b).

Training and testing data were provided to us from OCS as 207 sets of peer evaluations for four platoons, each containing three squads of roughly 18 candidates apiece. Peer evaluations are done at the squad level, and each set of inputs included about 17 scanned peer-evaluation sheets for each candidate. Sets contain an average of 1100 words and 100 to 200 sentences. Only seven of the 207 sets included a peer-evaluation summary form. Since the forms were handwritten, we had to extract the information by a time-consuming process of Microsoft Word dictation, transfer of the data to an Excel document for temporary storage, and loading the data into the relational database schema as in Figure 13. Dictation of one document set took approximately 20 minutes, so only a platoon and a half worth of documents were extracted to the database.

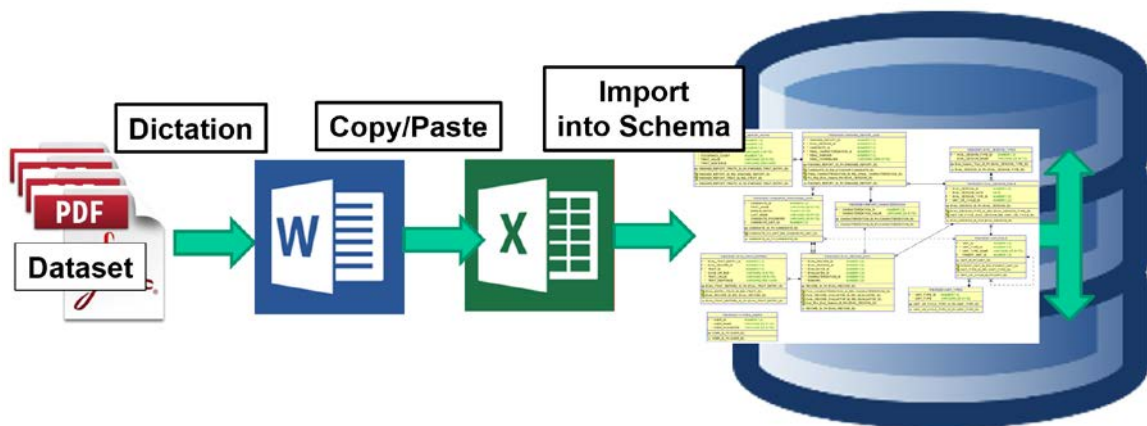


Figure 13. Thesis Extract, Transform and Load Workflow Diagram

## **B. DATA PREPARATION**

After connecting to the database and executing the SQL queries to get a squad's worth of data, several steps were needed to prepare the data for processing. We checked the spelling of all inputs, tokenized the input into sentences, stripped the punctuation, tokenized the sentences into words, removed proper possessive words, removed stop words and made N-grams, and finally stemmed or lemmatized the remaining N-grams.

Text-based probabilities were based on word and N-gram occurrences. To maintain text consistency, a single spell-checking algorithm is used at the beginning of preprocessing, both when training the model and when running on real-time data. To check spelling, we reviewed several Python packages (Figure 14). The "pattern.en" spellchecker is part of a larger suite of natural-language processing tools (Malik, 2020a). It was found to be the second fastest but made poor choices such as changing "At" to "It." The "spellchecker" package was slow and had trouble with misspellings such as "teh" (Barrus, 2018). Our choice, the "autocorrect" (Sondej, 2020) spellchecker, was the fastest of the three; it could run on words, N-grams, or whole sentences, and could be easily updated to recognize and correct task specific terms.

```

In [ ]: speller = Speller(lang='en')
print('autocorrect before adding context specific words:\t' + ' '.join([speller(word) for word in test.split()]))

speller = Speller(lang='en')
# Add military context specific words to Speller nlp_data dictionary with arbitrary frequency count of 100
special_words = ['SNC', 'snc', 'USMC', 'usmc', "SNC's", "snc's", 'OCS', 'ocs', 'NROTC', 'nrotc', 'TBS', 'tbs', 'MIDN',
for word in special_words:
    speller.nlp_data[word]=max([v for k,v in speller.nlp_data.items()])
speller.nlp_data['Of']=1
print('autocorrect after adding context specific words:\t' + ' '.join([speller(word) for word in test.split()]))

autocorrect before adding context specific words:      At OCS SN geos the distance even when sac's don't
autocorrect after adding context specific words:      At OCS SNC geos the distance even when snc's don't

In [ ]: test = "At OCS SNC geos teh distnaec evin whan snc's don't"
%timeit -n 100 [suggest(word)[0][0] for word in test.split()]
%timeit -n 100 [speller(word) for word in test.split()]
%timeit -n 100 [spellchecker.correction(word) for word in test.split()]

492 ms ± 32.6 ms per loop (mean ± std. dev. of 7 runs, 100 loops each)
167 ms ± 8.42 ms per loop (mean ± std. dev. of 7 runs, 100 loops each)
1.06 s ± 47.4 ms per loop (mean ± std. dev. of 7 runs, 100 loops each)

In [ ]: test = "At OCS SNC geos teh distnaec evin whan snc's don't"
print('pattern.en :\t', suggest(test)[0][0])
print('autocorrect :\t', speller(test))
print('speller :\t', spellchecker.correction(test))

pattern.en :      At OCS SNC geos teh distnaec evin whan snc's don't
autocorrect :      At OCS SNC geos the distance even whan snc's don't
speller :          At OCS SNC geos teh distnaec evin whan snc's don't

In [ ]: print('pattern.en :\t' + ' '.join([suggest(word)[0][0] for word in test.split()]))
print('autocorrect :\t' + ' '.join([speller(word) for word in test.split()]))
print('speller :\t' + ' '.join([spellchecker.correction(word) for word in test.split()]))

pattern.en :      It OCS SNC goes the distance even what sacks dont
autocorrect :      At OCS SNC geos the distance even whan snc's don't
speller :          At OCS SNC goes teh distance even what socks don't

```

Autocorrect: fastest

Autocorrect: only option to handle whole sentence input

Autocorrect: better capable to handle context specific text

Figure 14. Comparison of Spellchecking Packages Running in a Jupyter Notebook

We used the standard NLTK English stop words list, however, since peer-evaluation traits included a sentiment component, we removed “not” and any contractions using it (Figure 15) since they provide valuable information (Singh, 2019).

----- STOP WORDS -----					
i	me	my	myself	we	our
ours	ourselves	you	you're	you've	you'll
you'd	your	yours	yourself	yourselves	he
him	his	himself	she	she's	her
hers	herself	it	it's	its	itself
they	them	their	theirs	themselves	what
which	who	whom	this	that	that'll
these	those	am	is	are	was
were	be	been	being	have	has
had	having	do	does	did	doing
a	an	the	and	but	if
or	because	as	until	while	of
at	by	for	with	about	against
between	into	through	during	before	after
above	below	to	from	up	down
in	out	on	off	over	under
again	further	then	once	here	there
when	where	why	how	all	any
both	each	few	more	most	other
some	such	no	nor	<b>not</b>	only
own	same	so	than	too	very
s	t	can	will	just	<b>don</b>
<b>don't</b>	should	should've	now	d	ll
m	o	re	ve	y	<b>ain</b>
<b>aren</b>	<b>aren't</b>	<b>couldn</b>	<b>couldn't</b>	<b>didn</b>	<b>didn't</b>
<b>doesn</b>	<b>doesn't</b>	<b>hadn</b>	<b>hadn't</b>	<b>hasn</b>	<b>hasn't</b>
<b>haven</b>	<b>haven't</b>	<b>isn</b>	<b>isn't</b>	<b>ma</b>	<b>mightn</b>
<b>mightn't</b>	<b>mustn</b>	<b>mustn't</b>	<b>needn</b>	<b>needn't</b>	<b>shan</b>
<b>shan't</b>	<b>shouldn</b>	<b>shouldn't</b>	<b>wasn</b>	<b>wasn't</b>	<b>weren</b>

Figure 15. NLTK Stop Words with Words in Boldface to be Removed from List

We further decided to use only unigrams, bigrams, and trigrams of words since we did not have much training data. Preprocessed N-grams were stored in a pandas DataFrame object for processing. That let us quickly access preprocessed data later during experiments and final output generation.

### C. IDENTIFYING PREDICTORS

A key task was to identify clues (“predictors”) in our training data that help classify text, predict a report ranking or assessment, or identify patterns or anomalies in the data on a candidate. The Naïve Bayes classification technique benefits from careful choice of predictors, as do paraphrasing techniques.

Peer-evaluation data, collected from each candidate in a squad, should identify positive or negative traits, but the content of the final counseling paragraphs, written by squad and platoon staff, may reflect more than the six selected traits. The staff instead tries to provide a holistic assessment of performance metrics and ways to improve performance.

## **1. Correlation**

Squads vary in size, so to better compare data from multiple squads, we normalized final integer rankings by dividing by the number of candidates in the squad. Final assessments were assigned the values 1 for above the standard, 2 for within the standard, and 3 for below the standard. Using either the normalized final ranking or final assessment for each candidate as the response variable, and with the N-gram counts derived from the peer remarks for traits from each candidate as the predictor variables, we considered both the Pearson and Spearman correlation methods to rate the N-grams. Table 1 is a sample of the predictor and response variables for calculating our correlations between normalized ranking and number of occurrences of each N-gram in a candidate's peer remarks. We reasoned that finding a strong linear correlation between the number of occurrences of an N-gram and a high or low candidate ranking or assessment could identify N-grams useful in selecting sentences for paraphrasing or summarization, and the Pearson correlation would measure their strengths (Schober et al., 2018). Using the Python "pandas" package's built-in correlation function, we generated correlation values. Positive Pearson correlation values between the number of occurrences of an N-gram and the finalized ranking values meant that an N-gram was more closely correlated with a low-ranked candidate than a high-ranked candidate. As well, positive Pearson correlation values between the number of occurrences of an N-gram and the assessment value indicated that the N-gram was more closely correlated with a below-standards candidate. Conversely, negative Pearson correlation values for an N-gram when compared to either final ranking values or final assessment values meant that the N-gram was more closely correlated with either a high-ranked or above-standards candidate, respectively.

Table 1. Predictor and Response Variables from Training Data for Correlation by Normalized Final Rank

Response Variable	Predictor Variables					
	Rows correspond to candidates in the training set. Column headings are lemmatized N-grams from training data. Column values are occurrence counts of the N-gram for that candidate. Each N-gram column is correlated individually with the normalized rank column.					
Normalized rank	Confid	Knowledg	Alway help	Ever day	Work command presenc	Maintain good relat
1.000	0	0	0	0	0	0
1.000	0	8	0	0	0	0
1.000	2	0	0	0	1	0
0.944	1	5	1	0	0	0
0.938	7	0	0	0	0	1
0.938	1	8	0	0	0	0
0.889	2	0	1	0	2	1
0.875	3	1	1	0	0	1
0.875	4	5	0	0	0	0
0.833	2	0	1	1	0	1
0.813	0	0	0	0	0	0
0.813	1	0	0	0	0	0
0.813	2	0	0	0	0	1
...	...	...	...	...	...	...
0.611	0	0	4	0	1	0
0.563	0	0	0	0	0	0
0.563	1	0	0	0	0	1
0.556	5	0	0	1	1	0
0.500	0	7	0	0	0	1
0.500	2	2	0	0	0	0
0.500	4	0	0	0	0	0
0.500	1	8	0	0	0	0
0.444	1	0	0	0	1	1
0.438	3	0	0	0	0	1
...	...	...	...	...	...	...
0.167	1	0	0	1	0	0
0.125	0	6	0	0	0	0
0.111	0	7	0	0	0	0
0.063	2	2	0	0	0	0
0.063	0	12	1	0	0	0

Table 2 shows the correlation values for the sample N-gram predictors in Table 1. It is plain to see that the N-gram “Confid” occurs more in statements for candidates that have a high normalized-rank value in Table 1 and that is reflected in a stronger positive Pearson correlation value. The opposite can be said of the N-gram “Knowledg” which occurs more for candidates with low normalized-rank values. The other N-grams in Table 2 have very low Pearson and Spearman correlation values so that their occurrence counts do not strongly indicate anything.

Table 2. Correlation Values for Sample N-Grams from Table 1

Correlation Method and Response Variable	Predictor Variables					
	confid	knowledg	alway help	everi day	work command presenc	maintain good relat
<b>Pearson Correlation by Normalized rank</b>	0.357	-0.279	-0.05	-0.159	0.11	0.158
<b>Pearson Correlation by Assessment</b>	0.338	-0.174	-0.059	-0.079	0.127	0.175
<b>Spearman Correlation by Normalized rank</b>	0.366	-0.306	-0.051	-0.166	0.077	0.199
<b>Spearman Correlation by Assessment</b>	0.352	-0.193	-0.058	-0.079	0.091	0.233

Table 3 shows the top five unigrams, bigrams, and trigrams by Pearson correlation value using for both the high ranked and high performing candidates in our training set. The Spearman correlation identifies the strength of a monotonically increasing or decreasing relationship between ranks and can also help indicate if the N-gram will be a

good predictor. A low Spearman correlation value is a bad sign even if the Pearson value is relatively high. The Spearman correlation values are in Table 3 for comparison. In similar fashion, Table 4 shows the top five unigrams, bigrams, and trigrams by Pearson correlation values for the low ranked and low performing candidates in our training set with their corresponding Spearman correlation values for comparison. Most the Pearson correlation values were below 0.5 therefore no N-gram correlations were very strong. After calculating correlation values we filtered the resulting rated N-grams to derive the best set of the N-grams for predicting either high or low ranked candidates, and for predicting above or below standards candidates. We used a threshold of 1.5 standard deviations from the mean correlation value to filter. We used the Spearman correlation values to look for N-grams that were only weakly following a monotonic relationship and removed any that had a value of less than 0.2.

Table 3. Top Five Unigrams, Bigrams, and Trigrams Indicating Higher Ranked or Higher Performing Candidates Based on Pearson Correlation with Normalized Rank (Left) And Assessment (Right)

Normalized Rank			Assessment		
N-Grams Most Highly Correlated with Low Value (High Ranked) Candidates in Our Training Set	Pearson Correlation Value	Spearman Correlation Value	N-Grams Most Highly Correlated with Low Value (High Performing) Candidates in Our Training Set	Pearson Correlation Value	Spearman Correlation Value
	<b>Unigrams</b>			<b>Unigrams</b>	
Extens	-0.5123	-0.5154	Extens	-0.4137	-0.4220
Prior	-0.4424	-0.3892	Servic	-0.4026	-0.4310
Job	-0.4284	-0.4313	Prior	-0.3953	-0.3481
Experi	-0.4252	-0.4163	Experi	-0.3717	-0.3566
Appli	-0.422	-0.4206	Whole	-0.3415	-0.3428
<b>Bigrams</b>			<b>Bigrams</b>		
Extens knowledg	-0.4602	-0.4588	Get job	-0.4185	-0.4198
Get job	-0.4554	-0.4597	Prior servic	-0.4026	-0.4310
Prior servic	-0.4213	-0.4453	Extens knowledg	-0.3945	-0.3961
Knowledg prior	-0.4105	-0.4325	Knowledg prior	-0.3881	-0.3987
Sourc inform	-0.3991	-0.3957	Job done	-0.368	-0.3689
<b>Trigrams</b>			<b>Trigrams</b>		
Get job done	-0.3852	-0.3877	Get job done	-0.3680	-0.3689
Not alway display	-0.3635	-0.3713	Alway squar away	-0.3296	-0.3312
Alway squar away	-0.3467	-0.352	Reli get job	-0.3252	-0.3271
Reli get job	-0.3250	-0.327	Not alway display	-0.3055	-0.3270
Knowledg prior experi	-0.3175	-0.315	Take action situat	-0.2828	-0.2836

Corresponding Spearman correlation included for comparison

Table 4. Top Five Unigrams, Bigrams, and Trigrams Indicating Lower Ranked or Lower Performing Candidates Based on Pearson Correlation with Normalized Rank (Left) And Assessment (Right)

Normalized Rank			Assessment		
N-Grams Most Highly Correlated with High Normalized Rank Value (Low Ranked) Candidates in Our Training Set	Pearson Correlation Value	Spearman Correlation Value	N-grams Most Highly Correlated with High Assessment (Low Performing) Candidates in Our Training Set	Pearson Correlation Value	Spearman Correlation Value
	<b>Unigrams</b>				<b>Unigrams</b>
Wellb	0.3480	0.3528	pt	0.3542	0.3831
Confid	0.3572	0.3656	matter	0.3604	0.3375
Matter	0.3692	0.3477	pack	0.3734	0.3400
Within	0.4171	0.4189	within	0.3838	0.3828
Struggl	0.5054	0.4665	struggl	0.4721	0.4481
<b>Bigrams</b>			<b>Bigrams</b>		
Make quick	0.3385	0.3440	made big	0.3385	0.3364
Quick decis	0.3460	0.3408	make quick	0.3654	0.3899
Struggl pt	0.3590	0.3708	well other	0.3891	0.3881
Endur pt	0.3622	0.3681	endur pt	0.3951	0.3927
Well other	0.3858	0.3897	quick decis	0.3999	0.3788
<b>Trigrams</b>			<b>Trigrams</b>		
Everi pt event	0.3097	0.3185	get along everyon	0.2928	0.2596
Not always quick	0.3164	0.3154	not always quick	0.3385	0.3364
Work well other	0.3323	0.3352	work well other	0.3590	0.3576
Make quick decis	0.3349	0.3348	make quick decis	0.3698	0.3673
Physic endur pt	0.3622	0.3681	physic endur pt	0.3951	0.3927

Corresponding Spearman correlation included for comparison

## 2. Least Squares Regression

Linear regression fits a line to data. We chose least-squares linear regression model for this thesis, the most common form. It finds weights that, for each given set of N-grams and their frequencies, best predict the effect of the occurrence of each N-gram on an overall characterization of a sentence. We studied three characterizations: the normalized

candidate ranking, the candidate final assessment (1 being above standards, 2 being within standards, and 3 being below standards), and whether the statement matched a trait-sentiment pair.

#### **D. GENERATING THE FINAL COUNSELING DOCUMENT**

The goal of the system was to generate a counseling document that followed the OCS standard for peer-evaluation forms. Figure 16 is an example of the output we sought to generate. Areas highlighted in green summarize the final ranking (1) and assessment (2) for the candidate. Here we chose to use a standard opening statement template that we created based on review of available examples for all system counseling outputs. Areas highlighted in red contain trait-sentiment pair descriptions from peer-evaluation inputs. We extracted 10 to 20 peer-evaluation statements and selected three to represent each top trait-sentiment pair. We call this “trait-sentiment extraction.” Areas highlighted in yellow are a counseling paragraph created from a repository of counseling-paragraph statements partly from OCS and partly written by us. It uses whole statements extracted from this repository; six representing the candidate’s top six trait-sentiment pairs, and a closing sentence representing the candidate’s ranking and assessment. We call this “summary-statement extraction.”

CANDIDATE INTERVIEW FORM

<b>REASON FOR INTERVIEW</b>	PEER EVALUATION III	<b>1</b>	
<b>INTERVIEWER'S REMARKS</b>			
<p>The purpose of this counseling is to inform the candidate that he was ranked 16 of 16 in his squad during peer evaluation III. Furthermore, this counseling serves as a way for the candidate to identify both negative and positive trends in his performance and set goals to improve that performance.</p>			
<p><u>Negative Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Bearing.</b> SNC can sometimes appear visibly shaken or frustrated, especially under pressure. Also needs to work on his command presence in how he talks, acts, walks, and overall body language. Although committed to the task, and aware of the task, SNC has difficulty in executing it in a military manner. SNC was observed conducting a count in the field looking lost and confused and dazed from the training event. Carries self haphazardly, hunched, uniform and gear not straightened away. Shows defeated feelings clearly through voice, body language.</li> <li>- <b>Judgment.</b> Doesn't know or doesn't act on expectations of program and peers. Constantly surprising this candidate with new decisions that run contrary to baseline expectations. SNC frequently makes poor decisions for himself and how he manages his time. He often is disorganized which causes him to move slowly and be unprepared for deadlines. SNC's judgment is not 100% reliable or sound during SULE II. He could have made more sound decisions for the squad.</li> <li>- <b>Dependability.</b> SNC struggles to complete tasks in a timely manner or sometimes even at all. He has been witnessed to get frustrated with a situation and throw his hands up and quit. SNC's lack of preparation and slow speed make it hard to depend on him for tasks. It's hard to predict if he will pay close attention to detail or not. SNC's lack of judgment makes him an unreliable subordinate or peer to be relied on for performing a duty. Especially in tactical situations like SULE.</li> </ul> <p><u>Positive Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Courage.</b> SNC continues to fight and overcome setbacks he could have given up but he is still pushing getting better everyday. SNC does not appear perturbed by constant correction by staff and peers. Wigs into negative feedback. Sounds off usually. SNC will willingly place himself in uncomfortable or difficult situations to better himself and / or to aid his squad.</li> <li>- <b>Enthusiasm.</b> SNC seems committed to the course and becoming a Marine. Maximum effort is observed frequently. The SNC attacks every challenge head on. Since the last evaluation, this candidate has noticed SNC's constant enthusiasm and passion towards training. SNC has good enthusiasm for coming to Officer Candidates School and staying cheerful throughout the process.</li> <li>- <b>Tact.</b> SNC was able to respectfully answer questions about his faith when asked by other candidates and did not get angry when others asked him rude questions. SNC, even when being corrected, takes information like a grain of salt and tries to use that to his advantage. SNC interacts and deals with his peers effectively. He is polite and does not cause strife or create misunderstandings.</li> </ul>			
<p>SNC steps up when others are scared to do so. SNC is also advised to take assistance from their peers with a positive attitude. It appears as if SNC has failed to learn from his mistakes, as he continues to demonstrate below standard decision making ability. He has been witnessed to get frustrated with a situation and throw his hands up and quit. SNC's peers clearly see that he maintains a positive mental attitude. Candidates wish to be around SNC due to his personality. SNC needs to assess why they came to Officer Candidates School and ensure that they are here for the right reasons.</p>			
<b>CANDIDATE'S EXPLANATION (If more space is needed, continue comments on the back utilizing head-to-toe format)</b>			
PERFORMANCE ASSESSMENT:		<u>BELOW STANDARDS</u>	
CANDIDATE'S SIGNATURE		<b>2</b>	CAPT LAST, FI. MI.
<b>CANDIDATE INTERVIEWED (LAST NAME, FI. MI.)</b>	<b>COMPANY</b>	<b>PLATOON</b>	<b>DATE</b>
201, C	D	2nd	20200427

Figure 16. Example of Goal Output Counseling Document

## E. NATURAL-LANGUAGE PROCESSING: TRAIT-SENTIMENT EXTRACTION

For our thesis, we reduced the number of peer-evaluation statements for each top-six trait-sentiment pair from an average between 10–20 down to three statements that captured the most important information. A statement was generally a single sentence but could be as many as three. We decided that statements that contained the most frequent N-grams for that trait-sentiment pair in the training set were the most important to include because they captured information from the most peers. Term frequency rated and choose the best statement in the trait-sentiment descriptions for a given trait-sentiment pair. We also decided that most distinctive statements for a trait-sentiment pair were important to include. Information can be distinctive when compared to other statements about a candidate, or distinctive when compared to that of candidates in the squad. To choose the second kind of statement, we used inverse-document frequency on just the inputs for the candidate to find the rarest statement. We selected the third statement using inverse-document frequency calculations but applied to frequency in the entire set of statements for a trait-sentiment pair for the squad.

Trait-sentiment extraction in our system worked on data of one squad at a time. After we chose a squad to process, we calculated term-frequency information for each candidate and for the squad. Then we iterated over each trait-sentiment pair for each candidate in the squad, rating the statements by summing the N-gram values derived from our term-frequency calculations, and picked the top three sentences using the algorithm in Figure 17.

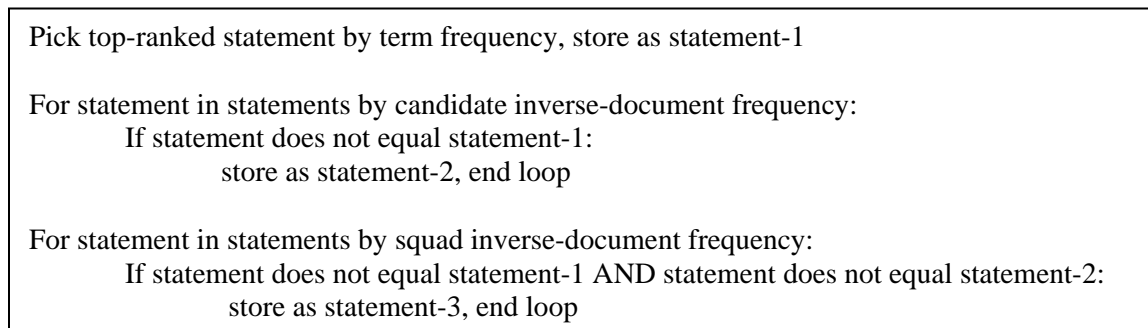


Figure 17. Extraction Method for Trait-Sentiment Statements

The rating system we used was based on (Malik, 2020b), though we also examined weighting as suggested by (Tiha, 2017). Three weighting methods we tested were not weighting the ratings at all, dividing statement ratings by the number of N-grams present, and penalizing sentences longer or shorter than the average length of the sentences in the group. For the third method we penalized statement ratings using formula (4.1).

$$\text{penalized\_rating} = \text{rating} * \left(1 - \left| \text{average\_length}_D - \frac{\text{len}(s_i)}{\text{average\_length}_D} \right| \right),$$

where:  $\text{len}(s_i)$  = number of N-grams in statement  $s_i$  of document  $D$ ,  
 $n$  = number of statements in document  $D$ ,

$$\text{average\_length}_D = \frac{\sum_{i=1}^n \text{len}(s_i)}{n}$$

(4.1)

## F. NATURAL-LANGUAGE PROCESSING: SUMMARY-STATEMENT EXTRACTION

To generate the final counseling paragraph of the peer-evaluation forms, we selected sentences from manually curated OCS examples with a few of our own additions. The final counseling paragraph needs a generic introductory sentence that gives statistics that went into the generation of the counseling (usually the count of above, within, and below standards assessments from peers), and a final candidate assessment and squad ranking. The body of the paragraph should mention the top three positive and top three negative traits as well as give advice on what the candidate has done well and on how the candidate can improve. The closing sentences, also generic in nature, should distinguish candidates based on the overall report assessment.

To do this, we created 32 bins of sentences; bins for each of 28 trait-sentiment pairs for paragraph-body sentences, a bin for each assessment characterization for closing sentences, and one bin for all opening sentences. We manually populated the bins with example sentences from the training set. Most examples came from OCS-provided peer-

evaluation forms, but since that was limited, we constructed some more based on our experience in training at OCS. We calculated TF-IDF values for all N-grams of a candidate for each top trait-sentiment-pair. Next, we chose a model for predicting rankings or assessments, either values for N-grams from correlations or weights for N-grams from regressions or both. Formula (4.2) calculates the sentence rating, the sum of a linear equation using information based on each N-gram in a sentence as input:

rating =

$$\sum_{i=1}^n \left[ \begin{array}{l} \text{TF-IDF}_{N\text{-gram}_i} * \\ (1 + \text{use\_correlation} * \text{adjustment} * \text{rank\_correlation\_value}_{N\text{-gram}_i}) * \\ (1 + \text{use\_correlation} * \text{adjustment} * \text{assessment\_correlation\_value}_{N\text{-gram}_i}) * \\ (1 + \text{use\_regression} * \text{regression\_weight}_{N\text{-gram}_i}) \end{array} \right],$$

where n is the number of N-grams in a given sentence,

use\_correlation is 1 if correlation model or both models are selected, 0 otherwise,

use\_regression is 1 if regression model or both models are selected, 0 otherwise,

and adjustment is 1 if candidate assessment is 3 or rank / number of candidates in the squad is greater than 2/3, -1 otherwise.

(4.2)

For the correlation model, since positive values correlated with poor performing candidates, we applied the “adjustment” to flip the sign on the values if the candidate was either an above-standards or within-standards performer or if they were ranked in the top two-thirds of the squad. The sentence with the highest rating was used in the summary paragraph.

Each OCS training unit is different, and writing styles change as staff rotates. It was therefore desirable that final counseling output options reflect the style of the users to some extent. To achieve this end, we desired to classify new sentences and place them in the correct bins automatically and manage bin size for continued long-term updating potential.

We developed a bootstrapping method to expand our training set and cull less useful examples. New examples were classified and assigned to a class bin. After training, each bin with more examples than the threshold size would have all sentences rated, and the lowest rated sentences were removed.

To classify sentences, we used a multinomial Naïve Bayes classifier. This assigned the sentence to the class that had the greatest rating based on its class-specific Naïve Bayes classifier. On training data of 126 potential sentences, we did term-frequency analysis on the N-grams, from which we constructed our Naïve Bayes classifiers. Several traits were rare, mostly those with positive sentiments, so we reduced the traits to seven in our experiments.

## **G. COMPILING A FINISHED DOCUMENT**

The last stage of processing assembled a summary peer-evaluation paragraph. Personal experience has shown it best to provide negative counseling recommendations between positive ones, so we provided one positive comment, followed by three negative comments, then two positive comments, to create the counseling paragraph. For the closing sentence, one sentence about the report type that matched the report characterization was selected using similar ranking techniques.

Figure 18 shows the overall system design we implemented to collect, analyze, process, and output peer-evaluation products. It uses the analytics, extraction, and compilation pipeline in Figure 19 which is the data processing pipeline that analyzes database inputs and generates the final document outputs for each candidate. It also includes a “commander’s dashboard” component that we were unable to implement and discuss our conclusion as future work. We trained the models on our training set (two squads from First Platoon, and one squad each from Second and Fourth Platoons) and tested them on our testing set (one squad from First Platoon). We also tested on candidates for which we were provided a final peer-evaluation form.

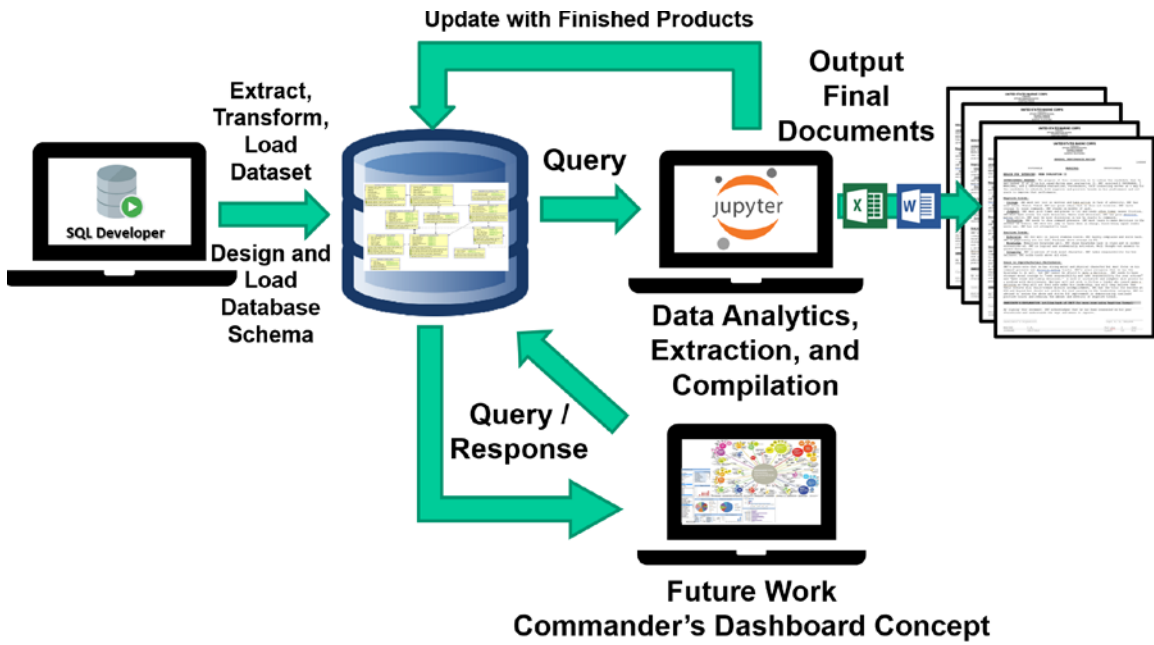


Figure 18. Thesis Peer-Evaluation System Overview

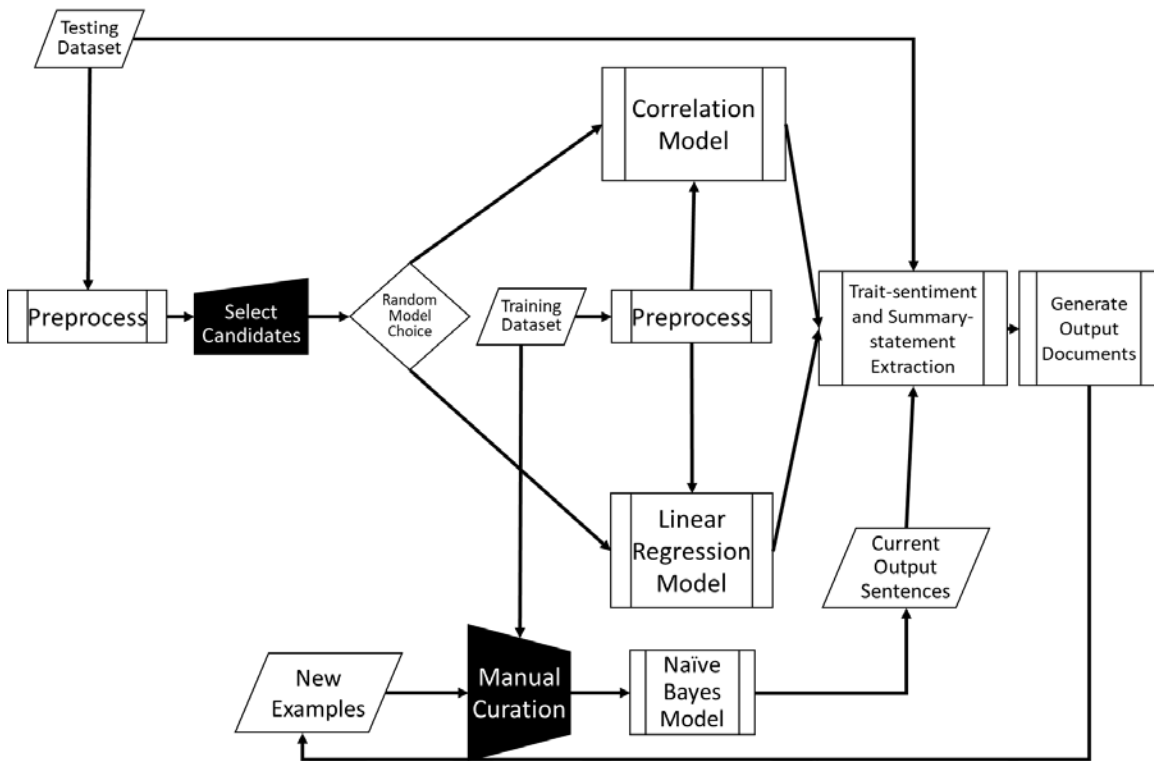


Figure 19. Analytics, Extraction, and Compilation Pipeline

In this pipeline, training and testing data is retrieved from the database and run through the preprocessing stages. Manual examination of the Naïve Bayes model training data is required since it does not come pre-labeled. Next, we select candidates to process; at OCS, a staff member would process a squad together. For testing in this thesis, we selected from the testing set: three above-standards performers, three within-standards performers, and three below-standards performers; all selected lacked final counseling outputs. Also, we selected one candidate whose data did include a final counseling. The candidates selected had varying final ranks. A model type was chosen at random to extract sentences for each candidate, and a final document was produced. SQL commands in a Jupyter Notebook aggregated data, Python code compiled the reports and wrote final report data to an Excel file. A Word template, connected to the Excel file using the mail-merge process, printed the final outputs in the OCS standard Word format.

THIS PAGE INTENTIONALLY LEFT BLANK

## V. RESULTS

### A. TRAIT-SENTIMENT EXTRACTION

Our attempts at trait-sentiment extraction were promising. In Table 5, data is shown for a candidate who demonstrated negative enthusiasm during the training period. Statement ratings in this table were not weighted. The terms “quiet” and “seems” appeared often in the eight inputs for this trait-sentiment pair which affected the rating for statements rating by term frequency only. A unique feature noticed by the squad was that the candidate was usually much less enthusiastic outside of a leadership billet; this concept was reflected by the high ratings for “overly” and “billet” in sentence two. Sentence three highlights a term distinctive when compared to those of the eight others in the squad with low enthusiasm: only this candidate was described with the term “lost,” whereas five of the eight were described with the term “quiet.”

Table 5. Trait-Sentiment Extraction Example without Weighting Statement Ratings

<b>Final trait-sentiment statement choices from those best-rated</b>		<ol style="list-style-type: none"> <li>1. SNC, due to his quiet nature, seems unenthusiastic.</li> <li>2. Outside of his billet, SNC does not seem overly enthusiastic.</li> <li>3. SNC is pretty quiet and can be easily lost amongst the platoons.</li> </ol>
<b>Statement rating method</b>	<b>Statement rating from summation of N-gram ratings</b>	<b>Top five statements from set of all statements from the peer-evaluation inputs for the trait-sentiment pair “enthusiasm-neg” for one candidate. Ranked by highest rating.</b>
<b>By TF:</b>	36 32 30 27 25	SNC, due to his quiet nature, seems unenthusiastic. Outside of his billet, SNC does not seem overly enthusiastic. At times SNC seems exhausted or uninterested in the task at hand. SNC is pretty quiet and can be easily lost amongst the platoons. SNC could display more energy and more of the daily task.
<b>By Candidate IDF:</b>	39.789 37.998 34.245 32.454 29.106	Outside of his billet, SNC does not seem overly enthusiastic. SNC, due to his quiet nature, seems unenthusiastic. SNC is pretty quiet and can be easily lost amongst the platoons. At times SNC seems exhausted or uninterested in the task at hand. SNC had shown lower enthusiasm for a few events in particular.
<b>By Squad IDF:</b>	32.636 31.096 27.449 26.111 22.485	SNC is pretty quiet and can be easily lost amongst the platoons. SNC, due to his quiet nature, seems unenthusiastic. At times SNC seems exhausted or uninterested in the task at hand. SNC had shown lower enthusiasm for a few events in particular. Outside of his billet, SNC does not seem overly enthusiastic.

As discussed in Chapter IV, we compared these results to those of two other methods to weight the statement rating calculations. The first divided each statement rating by the total number of N-grams in the statement. Doing so, we noticed significant changes in the top five statements ranked by term-frequency ratings when compared to those in Table 5, since ratings for shorter sentences became higher as in Table 6.

Table 6. Trait-Sentiment Extraction Example Weighting Statement Ratings by Number of N-Grams in Each Statement

<b>Final trait-sentiment statement choices from those best-rated</b>		<ol style="list-style-type: none"> <li>1. SNC could be louder at times.</li> <li>2. SNC had shown lower enthusiasm for a few events in particular.</li> <li>3. SNC is pretty quiet and can be easily lost amongst the platoons.</li> </ol>
<b>Statement rating method</b>	<b>Statement rating from summation of N-gram ratings</b>	<b>Top five statements from set of all statements from the peer-evaluation inputs for the trait-sentiment pair “enthusiasm-neg” for one candidate. Ranked by highest rating.</b>
<b>By TF:</b>	2.222 2.222 1.714 1.667 1.667	SNC could be louder at times. SNC is a little quiet at times. SNC, due to his quiet nature, seems unenthusiastic. At times SNC seems exhausted or uninterested in the task at hand. SNC could display more energy and more of the daily task.
<b>By Candidate IDF:</b>	1.940 1.903 1.895 1.809 1.803	SNC had shown lower enthusiasm for a few events in particular. SNC is pretty quiet and can be easily lost amongst the platoons. Outside of his billet, SNC does not seem overly enthusiastic. SNC, due to his quiet nature, seems unenthusiastic. At times SNC seems exhausted or uninterested in the task at hand.
<b>By Squad IDF:</b>	1.813 1.741 1.525 1.481 1.130	SNC is pretty quiet and can be easily lost amongst the platoons. SNC had shown lower enthusiasm for a few events in particular. At times SNC seems exhausted or uninterested in the task at hand. SNC, due to his quiet nature, seems unenthusiastic. SNC is a little quiet at times.

Table 7 shows the effect of penalizing statements that were further away from the average of the number of N-grams in all statements rated. The idea was that we would prefer a sentence that does not have too many or too few words compared to the average. Here we notice that most top five statements ranked by term-frequency or inverse-document-frequency rating are close to the same length.

Table 7. Trait-Sentiment Extraction Example Weighting Statements by Average Number of Words for All Ranked Statements

<b>Final trait-sentiment statement choices from those best-rated</b>		<ol style="list-style-type: none"> <li>1. At times SNC seems exhausted or uninterested in the task at hand.</li> <li>2. SNC is pretty quiet and can be easily lost amongst the platoons.</li> <li>3. SNC had shown lower enthusiasm for a few events in particular.</li> </ol>
<b>Statement rating method</b>	<b>Statement rating from summation of N-gram ratings</b>	<b>Top five statements from set of all statements from the peer-evaluation inputs for the trait-sentiment pair “enthusiasm-neg” for one candidate. Ranked by highest rating.</b>
<b>By TF:</b>	25.714 24.000 23.810 23.143 21.333	At times SNC seems exhausted or uninterested in the task at hand. SNC, due to his quiet nature, seems unenthusiastic. SNC could display more energy and more of the daily task. SNC is pretty quiet and can be easily lost amongst the platoons. Outside of his billet, SNC does not seem overly enthusiastic.
<b>By Candidate IDF:</b>	29.353 27.818 27.720 26.526 25.740	SNC is pretty quiet and can be easily lost amongst the platoons. At times SNC seems exhausted or uninterested in the task at hand. SNC had shown lower enthusiasm for a few events in particular. Outside of his billet, SNC does not seem overly enthusiastic. SNC could display more energy and more of the daily task.
<b>By Squad IDF:</b>	27.974 24.868 23.528 20.731 14.990	SNC is pretty quiet and can be easily lost amongst the platoons. SNC had shown lower enthusiasm for a few events in particular. At times SNC seems exhausted or uninterested in the task at hand. SNC, due to his quiet nature, seems unenthusiastic. Outside of his billet, SNC does not seem overly enthusiastic.

Each weighting method resulted in different trait-sentiment statement selections. Further qualitative analysis of the individual statements, beyond the scope of this thesis, is necessary to determine which method is best for this application. Our subsequent tests used the non-weighted method to compile our final documents.

## B. SUMMARY-STATEMENT EXTRACTION

Using the methods described in Chapter IV, summary statements extracted to create final counseling paragraphs met the requirements described but resulted in disjointed prose. They would still require editing by the staff member using the system. An example of our summary-sentence extraction is in Table 8 with the example provided from OCS for the same candidate. Notably, only one sentence in the OCS example was selected by our system to create the summary paragraph for the same candidate. Similarities are highlighted in yellow. Notably, the system selected mostly different sentences, even when

choosing from sentences that contained the exact sentences from the OCS provided summary.

Table 8. Comparison of the OCS Provided and System-Generated Summary Statement Paragraphs

Statement Derivation Source	Summary Statements Based on the Same Candidate from the Test Set
OCS Provided Example	<p>SNC needs to re-assess why he came to Officer Candidates School and ensure that he is here for the right reasons. SNC must consider the consequences of his actions and how his decisions have detracted from the overall success of his unit. <b>It appears as if SNC has failed to learn from his mistakes, as he continues to demonstrate below standards decision making ability.</b> SNC must take the lessons he is learning and be diligent in his application to his daily routine, in order to prevent making the same mistakes. By setting goals, SNC can focus on the specific tasks he struggles with and build good habits. SNC should seek out his squad mates to help him cultivate those new habits and perform to the standard held of all Officer Candidates. SNC is encouraged to take more time in the study of his Student Outlines and Candidate Regulations in order to reinforce the information he is learning. SNC is highly encouraged to seek out his fellow candidates and platoon staff for guidance in making improvements.</p>
System generated example using TF-IDF, correlation, and least squares regression model N-gram values to rate sentences.	<p>SNC steps up when others are scared to do so. SNC is also advised to take assistance from their peers with a positive attitude. <b>It appears as if SNC has failed to learn from his mistakes, as he continues to demonstrate below standards decision making ability.</b> He has been witnessed to get frustrated with a situation and throw his hands up and quit. SNC's peers clearly see that he maintains a positive mental attitude. Candidates wish to be around SNC due to his personality. SNC needs to assess why they came to Officer Candidates School and ensure that they are here for the right reasons.</p>

Yellow highlights indicate content shared between the summary statement paragraphs.

N-grams with higher TF-IDF values were most distinctive of the candidate. Comparing the summation of TF-IDF values for each sentence in the pool of available sentences for a given class helped identify the summary sentence that included information

most closely related to the candidate being evaluated. The TF-IDF values alone serve as a baseline for the individual candidate.

Statements for the summary paragraph were rated using formula 4.2 and may have included inputs from the correlation and linear regression models. We compare different methods of rating sentences in Table 9. When the effects of using one model or another on the rating derived using formula 4.2 are compared, we can more clearly see what each model highlights. Using the trained regression weights or correlation values for N-gram gave a global perspective. These models were trained on all top six trait-sentiment inputs, final ranks, and final assessments for all candidates in the training set, not just on the data for a single candidate. Comparing the sums of N-gram ratings derived from one of these models helped select good closing sentences. The least-squares regression model included encoded information for over 19000 N-grams and every trait-sentiment pair, opening sentence, and closing sentence, whereas the correlation model highlighted information for high-performing or low-performing candidates only and encoded information for about 3500 N-grams. Despite few encoded N-grams, the correlation model N-gram values were over one-hundred times larger on average than the least squares regression values and therefore made a greater impact. As expected, both models influenced the opening and closing sentences which were in turn influenced by final assessment of a candidate.

Table 9. Comparison of the Effects of Each Model on Summary-Statement Extraction for One Candidate

Statement Rating Method <sup>a</sup>	Summary Statements Extracted <sup>b</sup>
TF-IDF baseline	Candidate is the first to volunteer. SNC is significantly behind his peers in his ability to adapt to a military environment and still does not possess the discipline and bearing of an Officer Candidate. It appears as if SNC has failed to learn from his mistakes, as he continues to demonstrate below standards decision making ability in a constantly changing and uncertain environment. He is too insecure to lead and often just gets frustrated when things don't work out. SNC's peers clearly see that he maintains a positive mental attitude. SNC always maintains a positive attitude. SNC is also advised to seek assistance from their peers, to refer to their candidate regulations for information pertaining to standards of conduct, and to use the platoon staff for any necessary guidance in seeking improvement.
Least squares regression model weights only (use_correlation = 0, use_regression = 1)	SNC steps up when others are scared to do so. SNC can't be trusted to accomplish a wide range of tasks. It appears as if SNC has failed to learn from his mistakes, as he continues to demonstrate below standard decision making ability. SNC's major issues, as seen by their peers, circles around his performance as part of the platoon presented here at Officer Candidates School. SNC's peers clearly see that he maintains a positive mental attitude. Candidates wish to be around SNC due to his personality. SNC needs to assess why they came to Officer Candidates School and ensure that they are here for the right reasons.
Correlation model values only (use_correlation = 1, use_regression = 0)	SNC steps up when others are scared to do so. SNC is also advised to take assistance from their peers with a positive attitude. The easiest of tasks slow down SNC to a screeching halt. Furthermore, SNC is unable to 'read situations well' and made decisions that 'went against the orders of the leader' during the execution of SULE II. SNC's peers clearly see that he maintains a positive mental attitude. Candidates wish to be around SNC due to his personality. SNC needs to assess why they came to Officer Candidates School and ensure that they are here for the right reasons.
Correlation values, and least squares regression model weights combined; no TF-IDF (special case for this example). (use_correlation = 1, use_regression = 1)	SNC steps up when others are scared to do so. SNC can't be trusted to accomplish a wide range of tasks. It appears as if SNC has failed to learn from his mistakes, as he continues to demonstrate below standard decision making ability. SNC's major issues, as seen by their peers, circles around his performance as part of the platoon presented here at Officer Candidates School. SNC's peers clearly see that he maintains a positive mental attitude. Candidates wish to be around SNC due to his personality. SNC needs to assess why they came to Officer Candidates School and ensure that they are here for the right reasons.
TF-IDF, correlation, and least squares regression model weights combined. (use_correlation = 1, use_regression = 1)	SNC steps up when others are scared to do so. SNC is also advised to take assistance from their peers with a positive attitude. It appears as if SNC has failed to learn from his mistakes, as he continues to demonstrate below standard decision making ability. He has been witnessed to get frustrated with a situation and throw his hands up and quit. SNC's peers clearly see that he maintains a positive mental attitude. Candidates wish to be around SNC due to his personality. SNC needs to assess why they came to Officer Candidates School and ensure that they are here for the right reasons.

<sup>a</sup>Statements are derived by comparing ratings calculated using formula 4.2.

<sup>b</sup>Yellow highlights indicate that the sentence differs from its counterpart sentence in the respective location in the baseline TF-IDF output on the first row.

### C. SUMMARY-STATEMENT BIN UPDATES

Manually classifying final counseling sentences was a bit of a challenge. We labeled them based on the six trait-sentiment pairs for each counseling and the overall assessment stated in the counseling document. As discussed in Chapter IV, we generated prior probabilities for the initial 32 summary-statement classes in Table 10. Following the Bayesian methods described in Chapter III, we used term frequency to derive the likelihoods that individual N-grams were part of a class, which in turn let us calculate class conditional probabilities for a new summary statement. Examples of class conditional probabilities for three example phrases are in Table 11.

Table 10. Trait-Sentiment Prior Probabilities in Our Training Set for the 32 Classes

CLASS	EXAMPLES	PRIOR PROBABILITY	CLASS	EXAMPLES	PRIOR PROBABILITY
<b>opening-opening</b>	23	0.183	<b>integrity-pos</b>	0	0.0
<b>closing-above</b>	0	0.0	<b>integrity-neg</b>	0	0.0
<b>closing-within</b>	6	0.0476	<b>endurance-pos</b>	0	0.0
<b>closing-below</b>	8	0.0635	<b>endurance-neg</b>	8	0.0635
<b>justice-pos</b>	0	0.0	<b>bearing-pos</b>	0	0.0
<b>justice-neg</b>	0	0.0	<b>bearing-neg</b>	4	0.0317
<b>judgment-pos</b>	0	0.0	<b>unselfishness-pos</b>	0	0.0
<b>judgment-neg</b>	16	0.127	<b>unselfishness-neg</b>	3	0.0238
<b>decisiveness-pos</b>	0	0.0	<b>courage-pos</b>	0	0.0
<b>decisiveness-neg</b>	3	0.0238	<b>courage-neg</b>	0	0.0
<b>initiative-pos</b>	1	0.00794	<b>knowledge-pos</b>	1	0.00794
<b>initiative-neg</b>	11	0.0873	<b>knowledge-neg</b>	1	0.00794
<b>dependability-pos</b>	0	0.0	<b>loyalty-pos</b>	0	0.0
<b>dependability-neg</b>	33	0.262	<b>loyalty-neg</b>	0	0.0
<b>tact-pos</b>	0	0.0	<b>enthusiasm-pos</b>	0	0.0
<b>tact-neg</b>	2	0.0159	<b>enthusiasm-neg</b>	6	0.0476

Table 11. Class Conditional Probabilities Generated by Naïve Bayes Classifiers for Three Example Phrases, with Highest Values Highlighted

CLASS	“Study of his Student Outlines and Candidate Regulations”	“Centers around both his performance during physical training evolutions”	“Encouraged to seek out his fellow candidates and platoon staff”
opening-opening	-89.905	-112.302	-112.162
closing-above	-88.866	-111.181	-111.181
closing-within	-86.32	-113.617	-106.784
closing-below	-86.1	-113.847	-109.849
justice-pos	-88.866	-111.181	-111.181
justice-neg	-88.866	-111.181	-111.181
judgment-pos	-88.866	-111.181	-111.181
judgment-neg	-90.716	-113.176	-113.202
decisiveness-pos	-88.866	-111.181	-111.181
decisiveness-neg	-91.711	-114.197	-113.579
initiative-pos	-91.867	-110.7	-113.136
initiative-neg	-90.781	-113.268	-109.529
dependability-pos	-88.866	-111.181	-111.181
dependability-neg	-90.395	-112.264	-110.553
tact-pos	-88.866	-111.181	-111.181
tact-neg	-91.802	-114.289	-114.289
integrity-pos	-88.866	-111.181	-111.181
integrity-neg	-88.866	-111.181	-111.181
endurance-pos	-88.866	-111.181	-111.181
endurance-neg	-90.654	-108.125	-112.727
bearing-pos	-88.866	-111.181	-111.181
bearing-neg	-90.968	-114.11	-113.455
unselfishness-pos	-88.866	-111.181	-111.181
unselfishness-neg	-91.711	-113.528	-114.197
courage-pos	-88.866	-111.181	-111.181
courage-neg	-88.866	-111.181	-111.181
knowledge-pos	-91.867	-114.354	-114.354
knowledge-neg	-75.918	-114.354	-111.696
loyalty-pos	-88.866	-111.181	-111.181
loyalty-neg	-88.866	-111.181	-111.181
enthusiasm-pos	-88.866	-111.181	-111.181
enthusiasm-neg	-91.478	-113.965	-113.965

Our multinomial Naïve Bayes classifier, which compared the outputs of all 32 individual class Naïve Bayes classifiers, could correctly classify 91 of 126 sentences in the initial pool of OCS-provided summary statements into the 32 classes. The examples in Table 12 and Table 13 highlight two main issues. First, our second assumption in Chapter IV, that those writing evaluations would only discuss the six primary trait-sentiment pairs identified during the peer evaluation is not always true: in 33 of the 126 cases, the predicted trait was not one of the six associated with a candidate. Second, ambiguity in the sentence threw off the manual labeling.

Table 12. Issue with Manual Summary-Sentence Labeling: Summary Sentence was Never Associated with a Top-Six Trait-Sentiment Pair

EXAMPLE SUMMARY SENTENCES	ACTUAL AND PREDICTED LABELS	TOP SIX TRAIT-SENTIMENT PAIR LABELS ASSOCIATED WITH THE GIVEN SENTENCE	ISSUE
Candidate’s peers report that, SNC “carries himself haphazardly, hunched, uniform and gear not straightened away.”	ACTUAL : DEPENDABILITY-NEG PREDICTION: ENTHUSIASM-NEG	DEPENDABILITY-NEG JUDGMENT-NEG BEARING-NEG COURAGE-POS ENTHUSIASM-POS TACT-POS	Label prediction is likely correct, but it is not in top six trait-sentiment pairs.
SNC must consider the consequences of his actions and how his decisions have detracted from the overall success of his unit.	ACTUAL : INITIATIVE-NEG PREDICTION: JUDGMENT-NEG	INITIATIVE-NEG ENTHUSIASM-NEG ENDURANCE-NEG TACT-POS LOYALTY-POS JUSTICE-POS	Label prediction is likely correct, but it is not in top six trait-sentiment pairs.

Table 13. Issue with Manual Summary-Sentence Labeling: Summary-Sentence Content is Ambiguous

EXAMPLE SUMMARY SENTENCES	ACTUAL AND PREDICTED LABELS	TOP SIX TRAIT-SENTIMENT PAIR LABELS ASSOCIATED WITH THE GIVEN SENTENCE	ISSUE
SNC's actions take away from the squad's ability to prepare themselves for training because they have to constantly correct him.	ACTUAL : JUDGMENT-NEG PREDICTION: DEPENDABILITY-NEG	DEPENDABILITY-NEG JUDGMENT-NEG BEARING-NEG COURAGE-POS ENTHUSIASM-POS TACT-POS	The sentence belonged to more than one class.
SNC is also advised to seek assistance from their peers, to refer to their candidate regulations for information pertaining to standards of conduct, and to use the platoon staff for any necessary guidance in seeking improvement.	ACTUAL : CLOSING-WITHIN PREDICTION: CLOSING-BELOW	ENDURANCE-NEG BEARING-NEG DEPENDABILITY-NEG KNOWLEDGE-POS COURAGE-POS TACT-POS	Closing sentences are used for average and below average candidates.
SNC's deficiencies are negatively affecting his squad by forcing his peers to continuously supervise him at critical times.	ACTUAL : INITIATIVE-NEG PREDICTION: DEPENDABILITY-NEG	INITIATIVE-NEG ENTHUSIASM-NEG ENDURANCE-NEG TACT-POS LOYALTY-POS JUSTICE-POS	The sentence belonged to more than one class.

When we ran the classifier on potential summary sentences that we created based on experience at OCS, over 90% were misclassified for our 32 classes out of 55 new sentences with 21 sentiments correct and only three classes correctly predicted. Two issues were that new examples had N-grams not seen during training, and limited examples for some classes may have provided insufficient data for them. The first issue occurs when different OCS staff members have different word choices. We addressed the second issue by generating our own seven labels to reflect key themes in the counseling paragraphs: opening, closing, physical, team, communication, mental, and attitude. Extracting keywords, we got N-grams for each new class which we used for rating and labeling each sentence in the training set; sentence rating assigned one point for each word in a new sentence from a given class and the highest-class rating became the label. Table 14 summarizes these new classes and their probabilities.

Table 14. Trait-Sentiment Prior Probabilities for Each of Seven Locally Labeled Classes on OCS Provided Summary Sentences

<b>CLASS</b>	<b>EXAMPLES</b>	<b>PRIOR PROBABILITY</b>
<b>opening</b>	40	0.317
<b>closing</b>	16	0.127
<b>physical</b>	4	0.0317
<b>team</b>	26	0.206
<b>communication</b>	3	0.0238
<b>mental</b>	30	0.238
<b>attitude</b>	7	0.0556

Running a seven-class multinomial-Naïve-Bayes classifier for these new classes correctly classified in 124 of 126 sentences in the training set of OCS-provided summary statements and running it on our own created examples correctly classified about 50%. Some examples are in Table 15 and Table 16. We may need more new examples to train better.

Table 15. Outliers in 7-Classes Test on the OCS-Provided Summary Sentences

<b>OCS-PROVIDED SUMMARY SENTENCE</b>	<b>ACTUAL AND PREDICTED LABELS</b>	<b>ISSUE</b>
SNC is not reliable.	ACTUAL : OPENING PREDICTION : TEAM	This was part of longer opening set of quoted sentences. Since broken apart, label should be changed.
SNC's major issues, as seen by their peers, centers around his dependability and ability to contribute to the team effort as part of the platoon.	ACTUAL : TEAM PREDICTION : OPENING	Item was an opening sentence, but N-grams indicated to manual labeler that it should be labeled TEAM.

Table 16. Examples of Issues in 7-Classes Test on Locally Created Summary Sentences

TESTING DATA	ACTUAL AND PREDICTED LABELS	ISSUE
SNC is advised to review the above and strive for improvement in demonstrating continued positive traits and reducing the amount and severity of negative trends.	ACTUAL : CLOSING PREDICTION : TEAM	This is a common closing sentence but was misclassified each time because of its ambiguity.
Tact goes a long way with building bridges now that will pay dividends later on in SNC's career.	ACTUAL : COMMUNICATION PREDICTION : OPENING	The label seems reasonable, but several N-grams were not seen during training, resulting in misclassification.
SNC is mastering the ability to follow first, but at the same time has already started to lead where he can.	ACTUAL : TEAM PREDICTION : OPENING	Item was a body sentence, but N-grams indicated to manual labeler that it should be labeled OPENING.

#### D. SUMMARY

For trait-sentiment extraction, traditional term-frequency analysis sufficed and was powerful. Our approach to selecting important and unique sentences appeared successful and is easily duplicated for other multisource assessment and feedback systems. The models we trained to help summarization appeared to work well both independently and together to influence closing sentence choice. Also, the least-squares regression model, trained on peer-evaluation input data, may help create a better dataset for other algorithms to train on in the future. For detailed comparison, an example of an OCS final counseling document from our training set and the system-generated counseling document based on the same underlying data are provided in Appendix C.

## VI. CONCLUSION

The OCS peer-evaluation process and similar military processes can benefit from automation. A significant reduction in the man-hours spent aggregating, reviewing, analyzing, and synthesizing counseling documents from multi-source feedback is possible and desirable. Artificial-intelligence methods, trained on this data, could reduce overall processing time by automatically selecting useful remarks to help draft counseling documents under the supervision of unit staff. We developed a proof-of-concept system that generated draft counseling documents of reasonable quality. This ability to highlight key information for staff members at OCS would help them meaningfully interact with their candidates and could reduce unnecessary staff work.

The methods we studied were simple and could trace and explain their results. Building trust is critical to building ethical artificial intelligence to support personnel actions. However, the extractive nature of our summarization process did not result in counseling statements that stood on their own, and staff work is still necessary to edit them for clarity and context. As discussed in Appendix A, training artificial-intelligence systems on this current dataset is limited due to its size and scope. Further research in this area requires a richer dataset that includes both male and female candidates, staff rater output, and information from each OCS training program to eliminate any potential biases in the natural-language processing. Moving to a digital system now would help provide a richer data repository for analysis without overburdening the limited OCS staff. Furthermore, access to data over multiple training cycles enables research into trends and could help apply artificial-intelligence methods in other contexts such as a “commander’s dashboard” to see holistic trends within and between training units. Such research could provide tangible value to the institution and promote a fairer training, screening, and evaluation experience for officer candidates.

THIS PAGE INTENTIONALLY LEFT BLANK

## APPENDIX A. OCS DATA STATEMENT

The OCS peer-evaluation dataset is a collection of ranking and type characterization labels for 101 individual peer-evaluation input forms with labels for 606 peer-evaluation input form trait inputs as well as ranking, type characterization, trait labels for six traits, and a final counseling paragraph for seven final counseling outputs. This dataset represents most peer-evaluation inputs and outputs generated by an OCS training company for one peer-evaluation iteration during a spring 2020 training cycle.

A. CURATION RATIONALE To study the automatic paraphrasing and summarization of peer-evaluation data at Marine Corps entry level training, a mostly complete set of peer-evaluation input forms and final counseling outputs was collected from one OCS peer-evaluation iteration for one company during a Spring 2020 candidate cycle. Final outputs were human-written and served as examples of the current system understanding of paraphrasing and summarization quality expectations of peer-evaluation inputs. Training labels were derived directly from the peer-evaluation input forms in which candidates indicate which of 14 leadership traits were reflected by a peer in either a positive or negative way. Final counseling outputs are not labeled, so the researcher parsed them into sentences and assigned multiple labels from either the same 14 leadership traits with positive or negative sentiment or a secondary set of seven content labels derived from a holistic view of the 7 available final counseling outputs.

B. LANGUAGE VARIETY The data was collected in the form of emailed PDF files from the OCS Coordinator of Student Activities in summer 2020. Information about which varieties of English are represented was not available, but at least U.S. mainstream English is included from many regions of the United States.

C. SPEAKER DEMOGRAPHIC Speakers were not directly approached for inclusion in this dataset. This data statement was prepared based on information provided from the OCS Coordinator of Student Activities in summer 2020. Only demographic data for the OCS candidates (source of peer-evaluation inputs) was provided but not for the staff (source of final counseling outputs). Based on dataset curator military work experience at

OCS, the staff members were most likely either senior First Lieutenants or Captains in the Marine Corps, male, ranging in age from 25–32 years. Based on candidate data provided, speakers were all males, ranging in age from 22–35 years, with nearly two-thirds between 22–26 years. Race/Ethnicities by percentage were White 68%, Black 9%, Hispanic 13%, and Other 10%. Recruiting locations by percentage were 1st Marine Corps Recruiting District (MCRD) (Northeastern US) 19%, 4th MCRD (Eastern US) 16%, 6th MCRD (Southeast US) 11%, 8th MCRD (Southwest US) 13%, 9th MCRD (Midwest US) 10%, 12th MCRD (West US) 15% and other Marine Corps Recruiting Command locations 17%. No direct information is available about socioeconomic class. It was assumed that most speakers speak English as a native language. Age, race and region demographics were relatively evenly split between the four platoons.

D. ANNOTATOR DEMOGRAPHIC The dataset annotations were split between the 207 OCS candidates that filled out the peer-evaluation input forms and the researcher who parsed and annotated the final counseling outputs. Demographics for the OCS candidates were reviewed in section C. All candidates are taught how to properly fill out the peer-evaluation input forms, but no other information is available on whether any OCS candidates had other formal or informal training on peer-evaluations or expertise in a similar field. The researcher is a white, male, 35 years old, English as a native language, United States Marine Corps officer with 17 years of military experience, 3 of those spent working with Marine Corps Recruiting Command and 2 assignments to OCS as a platoon commander each for one PLC Juniors training cycle. The researcher has formal training in personnel evaluations and performance counseling and informal training in and experience conducting several types of peer evaluations.

E. SPEECH SITUATION Peer evaluations at OCS are done in a time-constrained and stressful environment. Candidates writing peer-evaluation inputs, and staff generating final counseling outputs represent a formal, synchronous, scripted, written language consisting of quantitative and qualitative data. The quantitative data consist of rankings and a peer performance assessment of above, with, or below average. Qualitative data consists of picking six traits reflected by the peer under evaluation and writing one to two sentences about how that candidate reflected the chosen trait. Candidates have limited time

for revisions. Candidates write all peer-evaluation inputs by hand in block letters and therefore tend to keep comments short and to the point. Platoon staff members writing final counseling outputs are also under time constraints to get feedback to their candidates and submit evaluations to their chain of command, and therefore tend to keep their remarks short, generally no more than one-half page typed. The intended audience for the peer-evaluation inputs are generally the peer on whom the input is written but also the platoon staff member. The intended audience for platoon staff is candidate being evaluated, but also the higher level staff who will use the formal counseling contents to make retention decisions. Of note, the dataset curator, namely the researcher, had to manually input all scanned document data into a local database. This task was done with the Microsoft Word dictation feature to read in three to four candidate input sets, about seventy scanned pages, and then using Microsoft spell checking and visual proofreading to validate the entered data. This process was time consuming and some errors during the transcription process likely occurred.

F. TEXT CHARACTERISTICS Most inputs and outputs are written in the third person referring to a candidate as Subject Named Candidate (SNC) or using the pronouns he/him. On occasion the proper noun Candidate followed by a last name is used in place of SNC. Also, on occasion the pronouns they/them are used in place of he/him. All inputs use the modality of text.

G. RECORDING QUALITY N/A

H. OTHER While it was the case that the data provided by OCS only included inputs and outputs from male officers and candidates, it should be understood that at least two companies per year will have female representation in both the inputs and outputs. Adding data to this dataset in the future will require a revision to this data statement.

I. PROVENANCE APPENDIX N/A

THIS PAGE INTENTIONALLY LEFT BLANK

## APPENDIX B. DATABASE SCHEMA

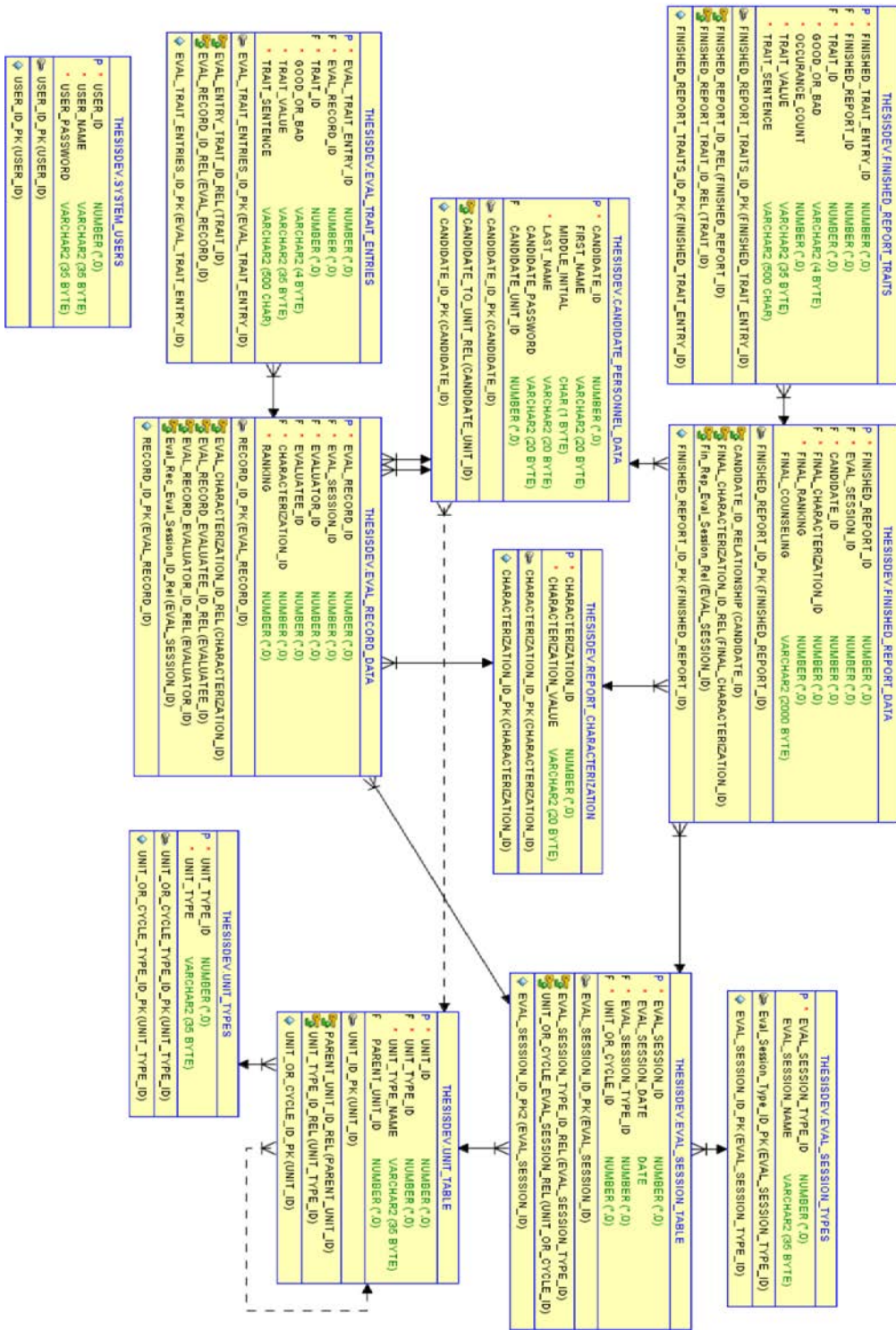


Figure 20. Database E-R Diagram

```

#Maj Andrew Nelson
#Sep 2020
#Thesis - Peer Eval System - Database Access Python Code
#All code run through jupyter notebook

# import for dataframe use
import numpy as np
import pandas as pd

# import for database access
import cx_Oracle

# For thesis work, must set and use unit ID values
# Unit ID values created during database schema upload
# Unit data for training
D11 = '64'
D12 = '65'
D21 = '67'
D42 = '74'

# Unit data for testing
D13 = '66'

# Currently selected unit
current_unit = D21

# primary SQL queries
candidate_count_query = "select count(candidate_id) from
CANDIDATE_PERSONNEL_DATA where CANDIDATE_UNIT_ID = "+current_unit

make_squad_view = "CREATE OR REPLACE VIEW current_session_info AS \
    SELECT a.eval_record_id, a.eval_session_id, a.evaluator_id, a.evaluatee_id,
a.characterization_id, a.ranking \
    FROM eval_record_data a, (SELECT eval_session_id FROM eval_session_table
WHERE unit_or_cycle_id = "+current_unit+" ) b \
    WHERE a.eval_session_id = b.eval_session_id"

make_squad_traits_view = "CREATE OR REPLACE VIEW current_session_traits AS \
    SELECT b.evaluatee_id, a.eval_record_id, a.pos_or_neg, a.trait_id,
a.trait_value, a.trait_sentence \
    FROM eval_trait_entries a, current_session_info b \
    WHERE a.eval_record_id = b.eval_record_id \
    ORDER BY a.eval_record_id, a.pos_or_neg desc"

# Generate SQL views for target squad session
with cx_Oracle.connect(username+'/' +password+'@'+host+':' +port) as con:
    cursor = con.cursor()
    num_candidates = cursor.execute(candidate_count_query)
    num_candidates = num_candidates.fetchone()[0]
    cursor.execute(make_squad_view)
    cursor.execute(make_squad_traits_view)

```

```

# Pull out raw inputs for target squad session into pandas DataFrames
# 1) Pull out eval session information (which session, rankings and
characterizations for all candidates in the session, etc)
# 2) Pull out eval session raw trait inputs

with cx_Oracle.connect(username+'/' +password+'@'+host+':' +port) as con:
    cursor = con.cursor()

    rs = cursor.execute("select * from current_session_info")
    col_names = [row [0] for row in cursor.description]
    type_rank_input_df = pd.DataFrame(data=rs, columns=col_names)

    rs = cursor.execute("select * from current_session_traits")
    col_names = [row [0] for row in cursor.description]
    traits_input_df = pd.DataFrame(data=rs, columns=col_names)

# After manipulating data, prepare and send it back to the database
# 1) Prepare and send back primary data for each report
# 2) Prepare and send back top trait outputs for each report

# Create SQL command as string for a report record output
def upload_finished_report_data(eval_session, candidate_id,
final_characterization_id, final_ranking, final_counseling):
    return "INSERT INTO FINISHED_REPORT_DATA (FINISHED_REPORT_ID,
EVAL_SESSION_ID, \
        CANDIDATE_ID, FINAL_CHARACTERIZATION_ID, FINAL_RANKING,
FINAL_COUNSELING) \
        VALUES ( FINISHED_REPORT_SEQ.nextval, "+str(eval_session)+,"
"+str(candidate_id)+," \
            "+str(final_characterization_id)+," "+str(final_ranking)+,"
'+final_counseling.replace("`", "`")+`)"

# Create SQL command as string for a trait output
def upload_finished_report_trait(finished_report_id, trait_id, pos_or_neg,
occurance_count, trait_value, trait_sentences):
    return "INSERT INTO FINISHED_REPORT_TRAITS (FINISHED_TRAIT_ENTRY_ID,
FINISHED_REPORT_ID, TRAIT_ID, POS_OR_NEG, \
        OCCURANCE_COUNT, TRAIT_VALUE, TRAIT_SENTENCES) \
        VALUES (
FINISHED_REPORT_TRAITS_SEQ.nextval, "+str(finished_report_id)+,"
"+str(trait_id)+," \
            "+str(pos_or_neg)+`", "+str(occurance_count)+,"
'+str(trait_value)+`", '+str(trait_sentences.replace("`", "`")+`)"

#upload finished report data
assess_ID_dict = {'ABOVE STANDARDS':1, 'WITHIN STANDARDS':2, 'BELOW
STANDARDS':3}
with cx_Oracle.connect(username+'/' +password+'@'+host+':' +port) as con:
    cursor = con.cursor()
    for candidate in mail_merge_df.index:
        # Execute SQL command to send data to database and commit changes
        cursor.execute(upload_finished_report_data(4, \
            mail_merge_df.loc [candidate, 'LNAME'], \
            assess_ID_dict [mail_merge_df.loc [candidate, 'FINAL ASSESSMENT']], \
            mail_merge_df.loc [candidate, 'FINAL RANK'], \
            mail_merge_df.loc [candidate, 'COUNSELING'])
        )
        cursor.execute("commit")

```

```

#upload finished report traits data
with cx_Oracle.connect(username+'/' +password+'@'+host+':' +port) as con:
    cursor = con.cursor()
    for trait in finished_report_traits_df.index:
        # Get finished_report_id to link to set of finished report traits
        rs = cursor.execute('select finished_report_id from finished_report_data
where candidate_id = '+str(finished_report_traits_df.loc [trait, 'evaluatee']))
        finished_report_id = rs.fetchone()[0]
        # Set variables for creating SQL string
        trait_id = str(finished_report_traits_df.loc [trait, 'trait_id'])
        pos_or_neg = finished_report_traits_df.loc [trait, 'pos_or_neg']
        occurrence_count = str(finished_report_traits_df.loc [trait,
'occurance_count'])
        trait_value = finished_report_traits_df.loc [trait, 'trait_value']
        trait_sentences = ' '.join(finished_report_traits_df.loc [trait,
'trait_sentences'])
        # Execute SQL command to send data to database and commit changes
        cursor.execute(upload_finished_report_trait(finished_report_id,
trait_id, pos_or_neg, \
                                                    occurrence_count,
trait_value, trait_sentences))
        cursor.execute("commit")

```

# APPENDIX C. COMPARISON OF FINAL COUNSELING STATEMENTS

## CANDIDATE INTERVIEW FORM

<b>REASON FOR INTERVIEW</b>	<b>PEER EVALUATION III</b>
<b>INTERVIEWER'S REMARKS</b>	
<p>The purpose of this counseling is to inform Candidate ██████ that the majority of the candidates in his squad ranked him below standards on Peer Evaluation III. SNC was marked below standard on 12 of 17 of his peer evaluations, identifying a trend of sub-standard performance when not in the spotlight or in view of platoon/company staff. Furthermore, this counseling serves as a way for him to identify his deficiencies in his performance and set goals to improve his future performance.</p> <p>SNC's major issues, as seen by his peers, are his lack of judgment eight times, dependability thirteen times, and bearing eight times. Specifically, Candidate ██████ peers report that, SNC "carries himself haphazardly, hunched, uniform and gear not straightened away. Shows defeated feelings clearly through voice, body language." "SNC struggles to complete tasks in a timely manner or sometimes, even at all. He has been witnesses to get frustrated with a situation and throw his hands up and quit." "SNC can't be trusted to accomplish a wide range of tasks. He is too insecure to lead and often just gets frustrated when things don't work out." "This candidate would not trust SNC to be physically ready for a real fire fight where there are not breaks." It is apparent that SNC is not a reliable member of his squad and can't be trusted to accomplish the mission. SNC is constantly unprepared for events and has trouble correctly performing basic daily tasks. SNC is significantly behind his peers in his ability to adapt to a military environment and still does not possess the discipline and bearing of an Officer Candidate. SNC's deficiencies are negatively affecting his squad by forcing his peers to continuously supervise him at critical times. SNC's actions take away from the squad's ability to prepare themselves for training because they have to constantly correct him. It appears as if SNC has failed to learn from his mistakes, as he continues to demonstrate below standards decision making ability in a constantly changing and uncertain environment.</p> <p>SNC needs to re-assess why he came to Officer Candidates School and ensure that he is here for the right reasons. SNC must consider the consequences of his actions and how his decisions have detracted from the overall success of his unit. It appears as if SNC has failed to learn from his mistakes, as he continues to demonstrate below standards decision making ability. SNC must take the lessons he is learning and be diligent in his application to his daily routine, in order to prevent making the same mistakes. By setting goals, SNC can focus on the specific tasks he struggles with and build good habits. SNC should seek out his squad mates to help him cultivate those new habits and perform to the standard held of all Officer Candidates. SNC is encouraged to take more time in the study of his Student Outlines and Candidate Regulations in order to reinforce the information he is learning. SNC is highly encouraged to seek out his fellow candidates and platoon staff for guidance in making improvements.</p>	
<b>CANDIDATE'S EXPLANATION (if more space is needed, continue comments on the back utilizing head-to-toe format)</b>	

ABOVE STANDARDS

WITHIN STANDARDS

BELOW STANDARDS

CANDIDATE INTERVIEWED (LAST NAME, FI, MI.)	COMPANY	PLATOON	DATE
██████████	D	2	██████████

1.2

Figure 21. Output Document from OCS Dataset

**CANDIDATE INTERVIEW FORM**

<b>REASON FOR INTERVIEW</b>	PEER EVALUATION III
<b>INTERVIEWER'S REMARKS</b>	
<p>The purpose of this counseling is to inform the candidate that he was ranked 16 of 16 in his squad during peer evaluation III. Furthermore, this counseling serves as a way for the candidate to identify both negative and positive trends in his performance and set goals to improve that performance.</p> <p><u>Negative Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Bearing.</b> SNC can sometimes appear visibly shaken or frustrated, especially under pressure. Also needs to work on his command presence in how he talks, acts, walks, and overall body language. Although committed to the task, and aware of the task, SNC has difficulty in executing it in a military manner. SNC was observed conducting a count in the field looking lost and confused and dazed from the training event. Carries self haphazardly, hunched, uniform and gear not straightened away. Shows defeated feelings clearly through voice, body language.</li> <li>- <b>Judgment.</b> Doesn't know or doesn't act on expectations of program and peers. Constantly suprising this candidate with new decisions that run contrary to baseline expectations. SNC frequently makes poor decisions for himself and how he manages his time. He often is disorganized which causes him to move slowly and be unprepared for deadlines. SNC's judgment is not 100% reliable or sound during SULE II. He could have made more sound decisions for the squad.</li> <li>- <b>Dependability.</b> SNC struggles to complete tasks in a timely manner or sometimes even at all. He has been witnessed to get frustrated with a situation and throw his hands up and quit. SNC's lack of preparation and slow speed make it hard to depend on him for tasks. It's hard to predict if he will pay close attention to detail or not. SNC's lack of judgment makes him an unrepliable subordinate or peer to be relied on for performing a duty. Especially in tactical situations like SULE.</li> </ul> <p><u>Positive Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Courage.</b> SNC continues to fight and overcome setbacks he could have given up but he is still pushing getting better everyday. SNC does not appear perturbed by constant correction by staff and peers. Wigs into negative feedback. Sounds off usually. SNC will willingly place himself in uncomfortable or difficult situations to better himself and / or to aid his squad.</li> <li>- <b>Enthusiasm.</b> SNC seems committed to the course and becoming a Marine. Maximum effort is observed frequently. The SNC attacks every challenge head on. Since the last evaluation, this candidate has noticed SNC's constant enthusiasm and passion towards training. SNC has good enthusiasm for coming to Officer Candidates School and staying cheerful throughout the process.</li> <li>- <b>Tact.</b> SNC was able to respectfully answer questions about his faith when asked by other candidates and did not get angry when others asked him rude questions. SNC, even when being corrected, takes information like a grain of salt and tries to use that to his advantage. SNC interacts and deals with his peers effectively. He is polite and does not cause strife or create misunderstandings.</li> </ul> <p>SNC steps up when others are scared to do so. SNC is also advised to take assistance from their peers with a positive attitude. It appears as if SNC has failed to learn from his mistakes, as he continues to demonstrate below standard decision making ability. He has been witnessed to get frustrated with a situation and throw his hands up and quit. SNC's peers clearly see that he maintains a positive mental attitude. Candidates wish to be around SNC due to his personality. SNC needs to assess why they came to Officer Candidates School and ensure that they are here for the right reasons.</p>	
<b>CANDIDATE'S EXPLANATION (If more space is needed, continue comments on the back utilizing head-to-toe format)</b>	

PERFORMANCE ASSESSMENT:

BELOW STANDARDS

\_\_\_\_\_  
CANDIDATE'S SIGNATURE

\_\_\_\_\_  
CAPT LAST, FI. MI.

CANDIDATE INTERVIEWED (LAST NAME, FI. MI.)	COMPANY	PLATOON	DATE
201, C	D	2nd	20200427

Figure 22. System Generated Output Document

## **APPENDIX D. THREE ABOVE, THREE WITHIN, AND THREE BELOW-STANDARDS SYSTEM-GENERATED COUNSELING OUTPUTS**

For the following documents, the last name is the candidate number assigned to the document in the testing set and the first name is either an “L” if the linear regression model weights impacted summary-sentence extraction or “C” if correlation values did. Some typos occurred and are most likely the result of errors occurring during dictation to digitize the dataset. Upon review of the following nine documents, we note that the first three share a closing sentence which is judged to be a sentence corresponding to an above-standards candidate. Likewise, the second three share a similar within-standards closing sentence. While the last three do share a below-standards closing sentence, only one of those candidates is assessed as below standards. However, due to the low final ranking of these candidates, the system decided to rate and select a below-standards closing sentence and we think that is the right choice.

## CANDIDATE INTERVIEW FORM

<b>REASON FOR INTERVIEW</b>	PEER EVALUATION III
<b>INTERVIEWER'S REMARKS</b>	
<p>The purpose of this counseling is to inform the candidate that he was ranked 1 of 16 in his squad during peer evaluation III. Furthermore, this counseling serves as a way for the candidate to identify both negative and positive trends in his performance and set goals to improve that performance.</p> <p><u>Negative Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Justice.</b> SNC sometimes shows slight favoritism to his own reteam. normally keeps to his own group, but is breaking out a bit more. SNC has improved upon his perceived favoritism, but still needs to make the effort to show impartiality.</li> <li>- <b>Tact.</b> candidate will sometimes make it known Is displeased if candidate didn't take advice. SNC has gotten better with his criticisms but could improve a bit more. SNC could get along better with team members.</li> <li>- <b>Enthusiasm.</b> candidate does not always display a sense of motivation. candidate professional demeanor doesn't reflect his interest to be here. can I couldn't look more enthusiastic during training events.</li> </ul> <p><u>Positive Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Knowledge.</b> SNC has vast knowledge and is willing to share that knowledge to make other candidates better. candidates prior service knowledge made him a major asset to Soul E2 preparation. SNC is incredibly well versed in anything related to the military.</li> <li>- <b>Judgment.</b> candidate is always able to make timely good decisions. SNC's superior knowledge and intellect allows him to quickly come to sound judgments. SNC made immediate decisions in SULE II.</li> <li>- <b>Decisiveness.</b> SNC is direct and doesn't waiver between decisions he chooses to make. candidate did not hesitate with his decisions in Snc E2 and does not do so in general. candidate makes sound timely decisions as leader.</li> </ul> <p>SNC's peers note his knowledge of USMC history and OCS academic requirements and look up to him as a mentor. SNC's peers remark that SNC unfairly assigns tasks when in a billet. Tact goes a long way with building bridges now that will pay dividends later on in SNC's career. Specifically, SNC's peers report that, 'SNC does not show any enthusiasm.' SNC weighs all his options before making a decision. SNC demonstrates initiative and decisiveness, traits that are imperative for a Marine leader to thrive in a chaotic environment. SNC's peers obviously trust and look up to him, therefore SNC should accept the leadership challenges afforded to him at OCS, not rest on his laurels, and continue leading from the front.</p>	
<b>CANDIDATE'S EXPLANATION (If more space is needed, continue comments on the back utilizing head-to-toe format)</b>	

PERFORMANCE ASSESSMENT:

**ABOVE STANDARDS**

\_\_\_\_\_  
CANDIDATE'S SIGNATURE

\_\_\_\_\_  
CAPT LAST, FI. MI.

CANDIDATE INTERVIEWED (LAST NAME, FI. MI.)	COMPANY	PLATOON	DATE
7, L ?	D	1st	20200427

## CANDIDATE INTERVIEW FORM

<b>REASON FOR INTERVIEW</b>	PEER EVALUATION III
<b>INTERVIEWER'S REMARKS</b>	
<p>The purpose of this counseling is to inform the candidate that he was ranked 3 of 16 in his squad during peer evaluation III. Furthermore, this counseling serves as a way for the candidate to identify both negative and positive trends in his performance and set goals to improve that performance.</p> <p><u>Negative Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Courage.</b> This candidate hasn't seen SNC push up a great challenge self on voluntarily and do it. SNC has been blending into the general population it could make a better effort to stick out more. SNC Good stand to speak up and speak his mind more frequently.</li> <li>- <b>Enthusiasm.</b> At times SNC seems exhausted or uninterested in the task at hand. SNC is pretty quiet and can be easily lost amongst the platoons. SNC had shown lower enthusiasm for a few events in particular.</li> <li>- <b>Judgment.</b> SNC at the point of friction can make decisions that poorly impact the other candidates. Specifically, in Snc E two. SNC took SULE in squad out into open field when enemy had snipers and IDF at the ready. During SULEII, SNC let it attack through a 200 meter open field on a fortified position when excellent cover was available.</li> </ul> <p><u>Positive Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Dependability.</b> SNC is effective at getting candidates together to complete a task. Once given a task you know this task will be done. SNC was easy to rely on as a team leader during SULE two. May sound timely decisions during SULE II.</li> <li>- <b>Unselfishness.</b> SNC is very considerate towards others and always looks out for their team members. SNC is a major team player is always willing to help anyone in the squad. SNC has helped this candidate numerous times.</li> <li>- <b>Knowledge.</b> SNC is very knowledgeable in all course working constantly is making sure he is studying. He hasn't answered any question which is why everyone in the platoon looks to him for knowledge. SNC has a great background knowledge set that is very helpful for the class / practical applications.</li> </ul> <p>SNC's changing demeanor from barracks to field confuses those other candidates that expect the same intensity and leadership they view on a daily basis. SNC is timid and never volunteers for any assignments. Specifically, SNC's peers report that, 'SNC does not show any enthusiasm.' The easiest of tasks slow down SNC to a screeching halt. SNC is noted for an unselfish attitude and desire to better others. SNC's peers note his knowledge of USMC history and OCS academic requirements and look up to him as a mentor. SNC's peers obviously trust and look up to him, therefore SNC should accept the leadership challenges afforded to him at OCS, not rest on his laurels, and continue leading from the front.</p>	
<b>CANDIDATE'S EXPLANATION (If more space is needed, continue comments on the back utilizing head-to-toe format)</b>	

PERFORMANCE ASSESSMENT:

**ABOVE STANDARDS**

\_\_\_\_\_  
CANDIDATE'S SIGNATURE

\_\_\_\_\_  
CAPT LAST, FI. MI.

CANDIDATE INTERVIEWED (LAST NAME, FI. MI.)	COMPANY	PLATOON	DATE
12, C ?	D	1st	20200427

## CANDIDATE INTERVIEW FORM

<b>REASON FOR INTERVIEW</b>	PEER EVALUATION III
<b>INTERVIEWER'S REMARKS</b>	
<p>The purpose of this counseling is to inform the candidate that he was ranked 3 of 16 in his squad during peer evaluation III. Furthermore, this counseling serves as a way for the candidate to identify both negative and positive trends in his performance and set goals to improve that performance.</p> <p><u>Negative Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Integrity.</b> It isn't that the candidate lacks integrity, this candidate just ranks those traits lower because they haven't been exhibited is noticeably as the other leadership traits.</li> <li>- <b>Enthusiasm.</b> SNC shows a lack of enthusiasm toward certain tasks. SNC could have more positive energy that the rest of the platoon can grow off of. SNC sometimes lacks excitement, especially during routine tasks.</li> <li>- <b>Bearing.</b> SNC has a severe lack of bearing and has been connected multiple times. SNC has gotten platoon in trouble on multiple occasions for a goofing off. SNC has lost his bearing several times in front of the instructors.</li> </ul> <p><u>Positive Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Knowledge.</b> SNC has a wealth of knowledge from prior service that he helps fellow candidates with. SNC has an abundance of knowledge and he spends a lot of the time sharing that with all his peers. SNC's prior service knowledge is helpful perspective during conversations.</li> <li>- <b>Dependability.</b> SNC is continuously able to complete his task on time. SNC can be relied upon to complete any job assigned to him. SNC is easy to rely on to complete tasks or assist others.</li> <li>- <b>Initiative.</b> SNC knows what he has to do and doesn't need to be followed up with. SNC stepped up early as a leader in the platoon. SNC does not need to be told what to do.</li> </ul> <p>SNC's peers note his knowledge of USMC history and OCS academic requirements and look up to him as a mentor. SNC cannot be trusted to accomplish tasks on his own. Specifically, SNC's peers report that, 'SNC does not show any enthusiasm.' SNC is also advised to take assistance from their peers with a positive attitude. SNC's changing demeanor from barracks to field confuses those other candidates that expect the same intensity and leadership they view on a daily basis. SNC's major issues, as seen by their peers, centers around his dependability and ability to contribute to the team effort as part of the platoon. SNC's peers obviously trust and look up to him, therefore SNC should accept the leadership challenges afforded to him at OCS, not rest on his laurels, and continue leading from the front.</p>	
<b>CANDIDATE'S EXPLANATION (If more space is needed, continue comments on the back utilizing head-to-toe format)</b>	

PERFORMANCE ASSESSMENT: **ABOVE STANDARDS**

\_\_\_\_\_  
CANDIDATE'S SIGNATURE

\_\_\_\_\_  
CAPT LAST, FI. MI.

CANDIDATE INTERVIEWED (LAST NAME, FI. MI.)	COMPANY	PLATOON	DATE
15, C ?	D	1st	20200427

## CANDIDATE INTERVIEW FORM

<b>REASON FOR INTERVIEW</b>	PEER EVALUATION III
<b>INTERVIEWER'S REMARKS</b>	
<p>The purpose of this counseling is to inform the candidate that he was ranked 6 of 16 in his squad during peer evaluation III. Furthermore, this counseling serves as a way for the candidate to identify both negative and positive trends in his performance and set goals to improve that performance.</p> <p><u>Negative Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Justice.</b> He will correct a candidate on something that SNC does as well such as smiling. SNC sometimes shows slight favoritism to those closest to him. As a candidate squad leader, did not task Pt's fairly.</li> <li>- <b>Judgment.</b> SNC sometimes conduct himself in a manner opposite to what he expects of other candidates, such as staying locked up. SNC has decided to get seconds at the chow hall instead of organizing his squad. SNC often is a little too zealous and goes overboard with the yelling.</li> <li>- <b>Bearing.</b> SNC can be caught laughing or joking around during inappropriate times. SNC is constantly seen smiling and taking some events here as a joke. SNC appears to always have a half smile in his expression.</li> </ul> <p><u>Positive Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Enthusiasm.</b> SNC has a ton of positive energy that will motivate anyone. He never has low energy and always picks up the squads spirits. Is always sounding off and can easily say he has in a way enjoyed this entire process.</li> <li>- <b>Initiative.</b> SNC would volunteer immediately for any pray apps for SULE events. SNC will kick off knowledge or start on tasks without anyone telling him to. SNC has been heard asking for advice on how to better his performance here.</li> <li>- <b>Dependability.</b> SNC has been a go-to person if things need to get done. Asks are done correctly and quickly. SNC consistently has shown that he can be relied upon to complete any task or responsibility delegated to him. SNC was a major asset during SULE 2 as he was easy to rely on if placed in important role.</li> </ul> <p>SNC's peers clearly see that he maintains a positive mental attitude. SNC's peers remark that SNC unfairly assigns tasks when in a billet. It appears as if SNC has failed to learn from his mistakes, as he continues to demonstrate below standard decision making ability. SNC is also advised to take assistance from their peers with a positive attitude. SNC's major issues, as seen by their peers, centers around his dependability and ability to contribute to the team effort as part of the platoon. SNC's changing demeanor from barracks to field confuses those other candidates that expect the same intensity and leadership they view on a daily basis. SNC must demonstrate that he is able to be trusted with tasks given to him and has a positive attitude toward training and is committed to the success of his peers in addition to himself.</p>	
<b>CANDIDATE'S EXPLANATION (If more space is needed, continue comments on the back utilizing head-to-toe format)</b>	

PERFORMANCE ASSESSMENT:                      WITHIN STANDARDS

\_\_\_\_\_  
CANDIDATE'S SIGNATURE

\_\_\_\_\_  
CAPT LAST, FI. MI.

CANDIDATE INTERVIEWED (LAST NAME, FI. MI.)	COMPANY	PLATOON	DATE
2, L ?	D	1st	20200427

## CANDIDATE INTERVIEW FORM

<b>REASON FOR INTERVIEW</b>	PEER EVALUATION III
<b>INTERVIEWER'S REMARKS</b>	
<p>The purpose of this counseling is to inform the candidate that he was ranked 7 of 16 in his squad during peer evaluation III. Furthermore, this counseling serves as a way for the candidate to identify both negative and positive trends in his performance and set goals to improve that performance.</p> <p><u>Negative Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Enthusiasm.</b> SNC does not seem extremely enthusiastic outside of his billing. SNC shows relatively moderate levels of motivation for tasks or events. SNC could show more excitement in being more vocal on a day-to-day basis.</li> <li>- <b>Bearing.</b> SNC due to being quiet can seemed to be not confident in himself. SNC laughs at jokes at inappropriate moments. SNC will get occasionally flustered when presented by platoon staff.</li> <li>- <b>Initiative.</b> SNC is great leader and should step up more outside of being a leader through billets. Outside of a billet, SNC tends to let work come to him. SNC Israeli scene outside of Billings taking initiative.</li> </ul> <p><u>Positive Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Endurance.</b> SNC consistently puts out during all Pt events. SNC has excellent cardio and performance with ease during Pt. SNC has the perfect track record with Pt events.</li> <li>- <b>Dependability.</b> SNC usually gets assigned task completed and is relied upon. SNC can be given a task and have no issue getting it done without supervision. SNC was a great squad leader. Always had accountability and let us know what was going on.</li> <li>- <b>Tact.</b> SNC is always respectful and easy to work with in any scenario. SNC speaks to candidates in a way that is not insightful. SNC talks to squad members in a point manner.</li> </ul> <p>SNC's peers look up to him for his prowess at all physical events. Specifically, SNC's peers report that, 'SNC does not show any enthusiasm.' SNC is also advised to take assistance from their peers with a positive attitude. SNC does not display the ability to make daily, sound judgment calls. SNC's changing demeanor from barracks to field confuses those other candidates that expect the same intensity and leadership they view on a daily basis. Candidates wish to be around SNC due to his personality. SNC must demonstrate that he is able to be trusted with tasks given to him and has a positive attitude toward training and is committed to the success of his peers in addition to himself.</p>	
<b>CANDIDATE'S EXPLANATION (If more space is needed, continue comments on the back utilizing head-to-toe format)</b>	

PERFORMANCE ASSESSMENT:                      **WITHIN STANDARDS**

\_\_\_\_\_  
CANDIDATE'S SIGNATURE

\_\_\_\_\_  
CAPT LAST, FI. MI.

CANDIDATE INTERVIEWED (LAST NAME, FI. MI.)	COMPANY	PLATOON	DATE
14, C ?	D	1st	20200427

## CANDIDATE INTERVIEW FORM

<b>REASON FOR INTERVIEW</b>	PEER EVALUATION III
<b>INTERVIEWER'S REMARKS</b>	
<p>The purpose of this counseling is to inform the candidate that he was ranked 8 of 16 in his squad during peer evaluation III. Furthermore, this counseling serves as a way for the candidate to identify both negative and positive trends in his performance and set goals to improve that performance.</p> <p><u>Negative Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Courage.</b> SNC doesn't always speak up about opinion to get a task done. SNC needs to develop a voice and presence within the book soon. SNC times to stay relatively quiet in Group discussions despite being very knowledgeable.</li> <li>- <b>Initiative.</b> SNC usually waits to get told what to do before starting a task. SNC black sometimes the ability to step up and fill a leadership void. SNC is a valuable asset to the squad, he should take initiative more often.</li> <li>- <b>Enthusiasm.</b> SNC is quiet and reserved. It's hard to gauge his enthusiasm. SNC could use more positive exuberance on a daily basis. he does not display much energy or excitement usually.</li> </ul> <p><u>Positive Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Tact.</b> Candidate maintains good relations with all other candidates in the platoon. he is very friendly with everyone and knows how to correct others constructively. Candidate is good at diffusing situations whatever they may be.</li> <li>- <b>Knowledge.</b> SNC is knowledgeable and does not hesitate to help peers fill knowledge gaps. candidate can always be relied upon to answer any questions with a sound answer. contributed helpful knowledge during the SULE II group discussions.</li> <li>- <b>Dependability.</b> Candidate finishes every task he is assigned timely and with reliable results. SNC does an make sure he's always squared away for the good of the platoon. SNC was easy to rely on during SULE II events and in billet roles to get things done.</li> </ul> <p>Candidates wish to be around SNC due to his personality. SNC is timid and never volunteers for any assignments. SNC does not display the ability to make daily, sound judgment calls. Specifically, SNC's peers report that, 'SNC does not show any enthusiasm.' SNC's peers note his knowledge of USMC history and OCS academic requirements and look up to him as a mentor. SNC's changing demeanor from barracks to field confuses those other candidates that expect the same intensity and leadership they view on a daily basis. SNC must demonstrate that he is able to be trusted with tasks given to him and has a positive attitude toward training and is committed to the success of his peers in addition to himself.</p>	
<b>CANDIDATE'S EXPLANATION (If more space is needed, continue comments on the back utilizing head-to-toe format)</b>	

PERFORMANCE ASSESSMENT:                      **WITHIN STANDARDS**

\_\_\_\_\_  
CANDIDATE'S SIGNATURE

\_\_\_\_\_  
CAPT LAST, FI. MI.

CANDIDATE INTERVIEWED (LAST NAME, FI. MI.)	COMPANY	PLATOON	DATE
8, L ?	D	1st	20200427

## CANDIDATE INTERVIEW FORM

<b>REASON FOR INTERVIEW</b>	PEER EVALUATION III
<b>INTERVIEWER'S REMARKS</b>	
<p>The purpose of this counseling is to inform the candidate that he was ranked 11 of 16 in his squad during peer evaluation III. Furthermore, this counseling serves as a way for the candidate to identify both negative and positive trends in his performance and set goals to improve that performance.</p> <p><u>Negative Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Courage.</b> SNC could offer up his opinion more and volunteer his ideas to the platoon. SNC needs to speak up and give input on his good ideas. SNC could be a little more open and outspoken than the currently is.</li> <li>- <b>Initiative.</b> This candidate doesn't see SNC imitate much but he does thing well once assigned to him. Did not hear SNC volunteering for roles in SULE 2 or pray app. SNC could go more out of his way to be proactive about seeking out tasks to accomplish, such as during morning cleanup.</li> <li>- <b>Enthusiasm.</b> SNC is reserved in his day to day actions and presence. SNC could be louder and more enthusiastic about his daily tasks. SNC comes across as quite droll at times.</li> </ul> <p><u>Positive Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Endurance.</b> SNC pushed through great physical pain during SULE in order to support his squad. SNC's leg started to give out during SULE II but he still consistently put out for each evaluation. SNC performs well physically and is almost always at the top amongst the squad.</li> <li>- <b>Loyalty.</b> SNC stayed committed to team leaders during SULE events. He wants to be an officer and a Marine to advance the Corps not himself. SNC is devoted to his peers and can be relied on for help.</li> <li>- <b>Dependability.</b> SNC can be relied on to complete assigned tasks. SNC is counted on always to have himself and his gear squared away. SNC is a reliable leader and a reliable team member.</li> </ul> <p>SNC's peers look up to him for his prowess at all physical events. SNC is timid and never volunteers for any assignments. SNC does not display the ability to make daily, sound judgment calls. Specifically, SNC's peers report that, 'SNC does not show any enthusiasm.' SNC is very supportive of candidate leadership. A good follower will make a good leader. SNC's changing demeanor from barracks to field confuses those other candidates that expect the same intensity and leadership they view on a daily basis. SNC needs to assess why they came to Officer Candidates School and ensure that they are here for the right reasons.</p>	
<b>CANDIDATE'S EXPLANATION (If more space is needed, continue comments on the back utilizing head-to-toe format)</b>	

PERFORMANCE ASSESSMENT:                    WITHIN STANDARDS

\_\_\_\_\_  
CANDIDATE'S SIGNATURE

\_\_\_\_\_  
CAPT LAST, FI. MI.

CANDIDATE INTERVIEWED (LAST NAME, FI. MI.)	COMPANY	PLATOON	DATE
6, L ?	D	1st	20200427

## CANDIDATE INTERVIEW FORM

<b>REASON FOR INTERVIEW</b>	PEER EVALUATION III
<b>INTERVIEWER'S REMARKS</b>	
<p>The purpose of this counseling is to inform the candidate that he was ranked 13 of 16 in his squad during peer evaluation III. Furthermore, this counseling serves as a way for the candidate to identify both negative and positive trends in his performance and set goals to improve that performance.</p> <p><u>Negative Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Tact.</b> SNC sometimes it's too loud compared to billet holders. SNC needs to work on his leadership style. Telling at people loses them. SNC can be unnecessarily argumentative and can be a source of discord and disharmony within the squad.</li> <li>- <b>Unselfishness.</b> when doing a squad chore, SNC will work on his own personal tasks. SNC can be a bit stubborn with his point of view and unwilling to compromise for the team. SNC is consistently one of the last squad members to enter head for night cleanup.</li> <li>- <b>Bearing.</b> SNC Rakes bearing often to joke or laugh. SNC appeared someone never starring his SULEII brief. SNC shows great bearing in training, but during downtime could be more focused.</li> </ul> <p><u>Positive Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Enthusiasm.</b> SNC maintains high motivation and energy in a majority of tasks and events. SNC threw himself into SULEII prep above all others in first SULE squad. SNC is always in a great mood and helps lift of fellow candidates.</li> <li>- <b>Endurance.</b> SNC is one of the more physically adept in the squad. SNC there's not any trouble completing all physical training events. he excels at Pt events and never falls out.</li> <li>- <b>Courage.</b> SNC is not afraid to speak up and speak his mind at any scenario. without hesitation, SNC steps up to any challenge and is often the first to volunteer when needed. he's not afraid to speak up in Group discussions.</li> </ul> <p>SNC's peers clearly see that he maintains a positive mental attitude. Tact goes a long way with building bridges now that will pay dividends later on in SNC's career. SNC's peers indicated that his response to professional corrections from his peers was usually argumentative or disregarded. SNC is also advised to take assistance from their peers with a positive attitude. SNC's peers look up to him for his prowess at all physical events. SNC steps up when others are scared to do so. SNC needs to assess why they came to Officer Candidates School and ensure that they are here for the right reasons.</p>	
<b>CANDIDATE'S EXPLANATION (If more space is needed, continue comments on the back utilizing head-to-toe format)</b>	

PERFORMANCE ASSESSMENT:                    **WITHIN STANDARDS**

\_\_\_\_\_  
CANDIDATE'S SIGNATURE

\_\_\_\_\_  
CAPT LAST, FI. MI.

CANDIDATE INTERVIEWED (LAST NAME, FI. MI.)	COMPANY	PLATOON	DATE
10, C ?	D	1st	20200427

## CANDIDATE INTERVIEW FORM

<b>REASON FOR INTERVIEW</b>	PEER EVALUATION III
<b>INTERVIEWER'S REMARKS</b>	
<p>The purpose of this counseling is to inform the candidate that he was ranked 16 of 16 in his squad during peer evaluation III. Furthermore, this counseling serves as a way for the candidate to identify both negative and positive trends in his performance and set goals to improve that performance.</p> <p><u>Negative Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Tact.</b> SNC can often seem condescending making situations worse. SNC complained of squad performance after his SULE II scenario when overall they did well. SNC does not always deal with conflict in an appropriate manner.</li> <li>- <b>Endurance.</b> SNC needs to continue improving during Pt, especially running. Longer physical events seem to take bigger toll on SNC than others. SNC is seriously under prepared for the level of cardio-based exercises at OCS.</li> <li>- <b>Unselfishness.</b> SNC usually does not go out of his way to help his peers out. SNC still has a difficult time being part of a team. This works jointly with judgment. SNC is often seen making decisions that better them self rather than others.</li> </ul> <p><u>Positive Trends:</u></p> <ul style="list-style-type: none"> <li>- <b>Decisiveness.</b> SNC make good, sound decisions during SULE II. Without hesitation, SNC is able to make those decisions without delay. SNC comes to quick sound decisions, then communicates them effectively.</li> <li>- <b>Knowledge.</b> SNC has a lot of knowledge about OCS as well as the material within OCS. SNC has outside knowledge that benefits the members of the squad. He has a wealth of practical knowledge about OCS from his previous experience here.</li> <li>- <b>Courage.</b> SNC is not afraid to speak up and speak his mind when he feels he can contribute. SNC attempted a single envelopment during his SULE II scenario. SNC is not afraid to speak up against candidates in the wrong or to offer solutions in friction.</li> </ul> <p>SNC demonstrates initiative and decisiveness, traits that are imperative for a Marine leader to thrive in a chaotic environment. Tact goes a long way with building bridges now that will pay dividends later on in SNC's career. Does not accept feedback, tries to justify mediocrity. SNC's peers indicated that his response to professional corrections from his peers was usually argumentative or disregarded. SNC's peers note his knowledge of USMC history and OCS academic requirements and look up to him as a mentor. SNC steps up when others are scared to do so. SNC needs to assess why they came to Officer Candidates School and ensure that they are here for the right reasons.</p>	
<b>CANDIDATE'S EXPLANATION (If more space is needed, continue comments on the back utilizing head-to-toe format)</b>	

PERFORMANCE ASSESSMENT: **BELOW STANDARDS**

\_\_\_\_\_  
CANDIDATE'S SIGNATURE

\_\_\_\_\_  
CAPT LAST, FI. MI.

CANDIDATE INTERVIEWED (LAST NAME, FI. MI.)	COMPANY	PLATOON	DATE
5, C ?	D	1st	20200427

## LIST OF REFERENCES

- Barrus, T. (2018). *Pyspellchecker* [Python]. <https://pyspellchecker.readthedocs.io/en/latest/>
- Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6, 587–604. [https://doi.org/10.1162/tacl\\_a\\_00041](https://doi.org/10.1162/tacl_a_00041)
- Berger, D. (2019). *38th Commandant's planning guidance*. United States Marine Corps. <https://www.marines.mil/News/Publications/MCPPEL/Electronic-Library-Display/Article/1907265/38th-commandants-planning-guidance/>
- Bird, S., Loper, E., & Klien, E. (2020). *Natural Language Toolkit—NLTK 3.5 documentation* [Python]. <http://www.nltk.org/>
- Chadwick, B. A. (2013). Evaluating leadership. *Marine Corps Gazette; Quantico*, 97(1), 29–31.
- Criley, B. R. (2006). Performance counseling. *Marine Corps Gazette; Quantico*, 90(9), 68–70.
- Department of Defense. (2019, February 12). *Summary of the 2018 Department of Defense artificial intelligence strategy: Harnessing AI to advance our security and prosperity*. <https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF>
- Department of the Army. (2018). *Prioritizing efforts-readiness and lethality (update 8) (Army Directive 2018-07-8)*. [https://armypubs.army.mil/epubs/DR\\_pubs/DR\\_a/pdf/web/ARN9642\\_AD2018\\_07\\_8\\_Final.pdf](https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN9642_AD2018_07_8_Final.pdf)
- Deshpande, M. (2017, September 24). *Text classification tutorial with Naive Bayes*. Python Machine Learning. <https://pythonmachinelearning.pro/text-classification-tutorial-with-naive-bayes/>
- Edson, J. J. (1985). Counseling craze criticized. *Marine Corps Gazette (Pre-1994); Quantico*, 69(3), 32–33.
- EduBirdie. (2019, June 25). *Citing sources: Quote vs paraphrase vs summary*. <https://edubirdie.com/blog/quote-vs-paraphrase-vs-summary>
- Freihat, A. A., Giunchiglia, F., & Dutta, B. (2013). Regular polysemy in WordNet and pattern based approach. *International Journal on Advances in Intelligent Systems*, 6(3), 15.

- Garrett, M. G. (1996). The confusion between performance evaluation and counseling. *Marine Corps Gazette; Quantico*, 80(2), 39–40.
- Hardison, C. M., Zaydman, M., Oluwatola, T. A., Saavedra, A. R., Bush, T., Peterson, H., & Straus, S. G. (2015). *360-degree assessments: Are they the right tool for the U.S. military?* (Report No. RR998). RAND Corporation. [https://www.rand.org/pubs/research\\_reports/RR998.html](https://www.rand.org/pubs/research_reports/RR998.html)
- HQMC [Headquarters United States Marine Corps]. (1986). *USMC User's Guide to Counseling (NAVMC 2795)*. <https://www.marines.mil/News/Publications/MCPEL/Electronic-Library-Display/Article/899627/navmc-2795/>
- HQMC [Headquarters United States Marine Corps]. (2000). *Marine Corps Individual Records Administration Manual (MCO P1070.12K W/CH 1)*. <https://www.marines.mil/News/Publications/MCPEL/Electronic-Library-Display/Article/899281/mco-p107012k-wch-1/>
- HQMC [Headquarters United States Marine Corps]. (2017). *Marine Leader Development (MCO 1500.61)*. <https://www.marines.mil/News/Publications/MCPEL/Electronic-Library-Display/Article/1261547/mco-150061-canx-mco-150058-and-navmc-dir-150058/>
- HQMC [Headquarters United States Marine Corps]. (2018). *Performance Evaluation System (MCO 1610.7A)*. <https://www.marines.mil/News/Publications/MCPEL/Electronic-Library-Display/Article/1513503/mco-16107a-cancels-mco-16107/>
- Lepsinger, R., & Lucia, A. D. (1997). 360° feedback and performance appraisal. *Training*, 34(9), 62–70. ProQuest Central; Social Science Premium Collection.
- Malik, U. (2020a). *Python for NLP: Introduction to the Pattern library*. Stack Abuse. <https://stackabuse.com/python-for-nlp-introduction-to-the-pattern-library/>
- Malik, U. (2020b). *Text summarization with NLTK in Python*. Stack Abuse. <https://stackabuse.com/text-summarization-with-nltk-in-python/>
- Manning, C., Raghavan, P., & Schütze, H. (2009). *Introduction to information retrieval*. Cambridge University Press.
- McAninch, Col. K., U. S. Army. (2016). How the Army's multi-source assessment and feedback program could become a catalyst for leader development. *Military Review*, 96(5), 84–93.
- McCallum, A., & Nigam, K. (1998). A Comparison of event models for Naive Bayes text classification. *IN AAAI-98 WORKSHOP ON LEARNING FOR TEXT CATEGORIZATION*, 41--48.

- McCauley, C. D., Pollman, V., Bracken, D., Dalton, M., & Jako, R. (1997). *Should 360-degree feedback be used only for developmental purposes?* ProQuest Ebook Central. <https://ebookcentral.proquest.com>
- Navy Personnel Command. (2019, April 15). *Performance evaluation transformation*. <https://www.public.navy.mil/bupers-npc/career/performanceevaluation/Pages/transformation.aspx>
- Nelson, A. (2019, August 16). Do not overthink performance counseling, just get it done. *USNI Blog*. <https://blog.usni.org/posts/2019/08/16/do-not-overthink-performance-counseling-just-get-it-done>
- Niedziocha, C. (2014). Multirater assessments and the Marine Corps. *Marine Corps Gazette*, 98(9), 46–49.
- Nikhath, A. K., & Subrahmanyam, K. (2019). Feature selection, optimization and clustering strategies of text documents. *International Journal of Electrical and Computer Engineering (IJECE)*, 9(2), 1313–1320. <https://doi.org/10.11591/ijece.v9i2.pp1313-1320>
- Oracle Corporation. (2020a, September 2). *What is a relational database?* <https://www.oracle.com/database/what-is-a-relational-database/>
- Oracle Corporation. (2020b, September 3). *Welcome to cx\_Oracle's documentation*. Cx\_Oracle - Python Interface for Oracle Database. [https://oracle.github.io/python-cx\\_Oracle/](https://oracle.github.io/python-cx_Oracle/)
- Personnel Command Public Affairs. (2018, December 7). New performance evaluation tool tested by fleet sailors. *CHIPS The Department of Defense Technology Magazine*. <https://www.doncio.navy.mil/CHIPS/ArticleDetails.aspx?id=11026>
- Porter, M. (2001, October). *Snowball: A language for stemming algorithms*. <http://snowball.tartarus.org/texts/introduction.html>
- Princeton University. (2010). *WordNet: A lexical database for english*. <https://wordnet.princeton.edu/>
- Raschka, S. (2017). *Naive Bayes and text classification I - introduction and theory*. <http://arxiv.org/abs/1410.5329>
- SAS Institute Inc. (2020). *What is natural language processing?* [https://www.sas.com/en\\_us/insights/analytics/what-is-natural-language-processing-nlp.html](https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html)
- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5), 1763–1768. <https://doi.org/10.1213/ANE.0000000000002864>

- Singh, G. (2019, November 14). *Why you should avoid removing STOPWORDS*. Medium. <https://towardsdatascience.com/why-you-should-avoid-removing-stopwords-aa7a353d2a52>
- Sondej, F. (2020). *Autocorrect* [Python]. <https://github.com/fsondej/autocorrect>
- Tiha, A. (2017). *Survey of text summarization techniques*. <https://doi.org/10.13140/RG.2.2.26904.24324>
- Tirona, M., & Gislason, M. (2011, July 26). *The good, the bad, and the ugly of 360° evaluations*. Nonprofit Quarterly. <https://nonprofitquarterly.org/the-good-the-bad-and-the-ugly-of-360d-evaluations/>
- Training Command. (2020). *MCECST Manual—MCRD-LDR-1015*. [Handbook]. <https://www.trngcmd.marines.mil/Units/South-Atlantic/FMTB-E/Students/Student-Materials/MCECST-Manual/>
- Vaishnavi, V., Saritha, M., & Milton, R. S. (2013). Paraphrase identification in short texts using grammar patterns. *2013 International Conference on Recent Trends in Information Technology (ICRTIT)*, 472–477. <https://doi.org/10.1109/ICRTIT.2013.6844249>
- Wike, N. (n.d.). It's time to rethink 360 degree reviews. *The Military Leader*. Retrieved August 5, 2020, from <https://themilitaryleader.com/rethink-360-degree-reviews-guest-post/>

## **INITIAL DISTRIBUTION LIST**

1. Defense Technical Information Center  
Ft. Belvoir, Virginia
2. Dudley Knox Library  
Naval Postgraduate School  
Monterey, California