INSTITUTE FOR DEFENSE ANALYSES

Forecasting DAU Course Demand for the 4th Estate

David E. Hunter, Project Leader
J. M. Breuer
Margaret L. Giles
Kristen M. Guerrera
John J. Kane
EunRae Oh
Brian Q. Rieksts
David M. Tate
B. Douglas Wake
Kevin Y. Wu

April 2018
Approved for public release;
distribution is unlimited.
IDA Paper P-8788
Log: H 17-000589

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

About this Publication

This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-14-D-0001, FJ-7-4212, "Demand Forecasting Model for Acquisition Professionals," for the Office of the Secretary of Defense (OSD) and the 4th Estate Director, Acquisition Career Management (DACM). The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments

Thank you to Terry C. Heuring, Stanley A. Horowitz, and Paul M. Thompson for performing technical review of this document.

For More Information:

David E. Hunter, Project Leader
dhunter@ida.org, (703) 575-4686

David J. Nicholls, Director, Cost Analysis and Research Division
dnicholl@ida.org, (703) 575-4991

INSTITUTE FOR DEFENSE ANALYSES

IDA Paper P-8788

Forecasting DAU Course Demand for the 4th Estate

David E. Hunter, Project Leader
J. M. Breuer
Margaret L. Giles
Kristen M. Guerrera
John J. Kane
EunRae Oh
Brian Q. Rieksts
David M. Tate
B. Douglas Wake
Kevin Y. Wu

Executive Summary

The 4th Estate Director, Acquisition Career Management (DACM) represents civilians assigned to the defense agencies outside the military departments and combatant commands. His responsibilities include collaborating with the defense agencies on all aspects of career development. As part of his duties, the DACM would greatly benefit from having an improved ability to forecast future demand for training courses.

The 4th Estate DACM asked the Institute for Defense Analyses (IDA) to develop a model to forecast demand from civilians in 4th Estate agencies for Defense Acquisition University (DAU) training courses. This is the final publication from the IDA project.

Background

The defense acquisition workforce currently contains approximately 150,000 professionals, of whom nearly 27,000 are civilians in the 4th Estate defense agencies. Each of these individuals is required to become certified in accordance with standards established following the enactment of the Defense Acquisition Workforce Improvement Act (DAWIA).

Certification standards are published for each career field and level (e.g., Contracting Level II). The standards may include minimum education and experience requirements. The standards also identify all training courses required for DAWIA certification. Each acquisition position is designated a career field/level combination. Certification is not required for being hired into position. Once in an acquisition position, however, individuals have 24 months to become certified.

The acquisition and functional training certification requirements are the focus of this research paper. In particular, we are interested in forecasting demand for the training courses that require in-person attendance at DAU. These training courses are offered throughout the year at various DAU locations according to a published schedule.

The 4th Estate leadership is required early each year to submit forecasts for the number of course seats that will be needed by the 4th Estate workforce. This forecast, when combined with the forecasted demand from the Services, serves as an input to DAU when scheduling courses.

Data and Methodology

We collected and analyzed detailed, historical time series data from the last 10 years on the 4th Estate workforce and observable training course interactions. We added variables to these datasets and constructed an estimate of historical demand from the observable individual interactions with each training course. We then employed a combination of modern machine learning and more traditional predictive analytic techniques to construct a demand forecast for each resident training course included in our research.

Results

IDA developed a predictive analytical model to forecast demand for each of the 63 most popular courses. We used this model to compute forecasts of demand for acquisition training classes in fiscal year (FY) 2019 from 4th Estate agencies. These forecasts are shown in this paper and also in the computer model that was delivered to the sponsor.

We tested the methodology using actual data and forecasts for FY 2015 and FY 2016 and found that the IDA-developed forecast achieved a 49 percent improvement in accuracy over past 4th Estate forecasts. The IDA forecast was notably closer to the actuals than the legacy 4th Estate forecast for 50 of the 63 courses. For the courses for which the IDA forecast was better, the improvement averaged 100 seats closer to the actual demand.

IDA also developed a user-friendly computer model. A key feature of this IDA-developed tool is the ability to easily perform sensitivity analyses and estimate the effects of any changes (e.g., travel bans, hiring freezes, workforce expansions) on overall demand.

With a more accurate course demand forecast, the 4th Estate DACM is better positioned to ensure access to desired training courses for the 4th Estate acquisition workforce.

Contents

1. Introduction

The 4th Estate Director, Acquisition Career Management (DACM) asked the Institute for Defense Analyses (IDA) to develop a model to forecast the demand for training courses from civilian acquisition professionals in the Department of Defense (DoD) 4th Estate agencies. This is the final publication from the IDA project.

A. Background

The DoD 4th Estate consists of all organizational entities in DoD that are not in the military departments or combatant commands. Currently, there are approximately 27,000 acquisition professionals in 4th Estate agencies.

The 4th Estate DACM represents civilians assigned to the 4th Estate. His duties include collaboration with the defense agencies on all facets of career development and management of the defense acquisition workforce. Part of his responsibility is to ensure access to training courses required for certification in accordance with the Defense Acquisition Workforce Improvement Act (DAWIA) of 1990.

The major goal of DAWIA was to professionalize the DoD acquisition workforce, which currently contains approximately 150,000 members across the Army, Navy, Air Force, and 4th Estate. Certification standards were established that identify the training, education, and experience required for certification. These certification standards depend on two factors, the career field and career level assigned to each acquisition position.

The career field, as the name implies, describes the focus area of the acquisition position. All acquisition positions in DoD are assigned to one of the following 14 career fields:

- Auditing
- Business – Cost Estimating
- Business – Financial Management
- Contracting
- Engineering
- Facilities Engineering
- Industrial Contract Property Management
- Information Technology

- Life Cycle Logistics
- Production, Quality, and Manufacturing
- Program Management
- Purchasing
- Science and Technology Manager
- Test & Evaluation

Additionally, each acquisition position is assigned a level, which characterizes the seniority of the position. Each of the acquisition positions is assigned to one of the following three levels:

- Level I: Basic or Entry Level (GS 05–08)
- Level II: Intermediate or Journeyman Level (GS 09–12)
- Level III: Advanced or Senior Level (GS 13 and above)

Certification standards have been established for every career field/level combination. For example, Figure 1 shows the core certification standards for a Level II position in the Contracting career field. Entries in the Acquisition Training and Functional Training sections indicate required Defense Acquisition University (DAU) training courses.

Core Certification Standards (required for DAWIA certification)	
Acquisition Training	**ACQ 101** Fundamentals of Systems Acquisition Management
Functional Training	● **CON 200** Business Decisions for Contracting ● **CON 216** Legal Considerations in Contracting ● **CON 270** Intermediate Cost and Price Analysis **(R)** ● **CON 280** Source Selection and Administration of Service Contracts **(R)** ● **CON 290** Contract Administration and Negotiation Techniques in a Supply Environment **(R)** ● **CLC 051** Managing Government Property in the Possession of Contractors ● **CLC 056** Analyzing Contract Costs ● **HBS 428** Negotiating
Education	● At least 24 semester hours in accounting, law, business, finance, contracts, purchasing, economics, industrial management, marketing, quantitative methods, or organization and management ● Baccalaureate degree (Any Field of Study)
Experience	2 years of contracting experience.

Figure 1. Certification Standards for Level II Contracting

All personnel in acquisition positions have 24 months to achieve the certification standards for their position. The certification requirement cannot be waived. At times, however, personnel may be given an extension from the 24-month deadline.

B. Training Courses

Acquisition and functional training courses are a major component of the requirements for certification. A list of training courses that need to be completed for certification is provided for every career field/level combination. Some required courses are available online. Other courses, referred to as resident courses, require in-person attendance. The focus of this paper is forecasting demand for these resident courses.

DAU training courses are offered throughout the year at various locations in the United States and abroad according to a published schedule. A course may be offered multiple times throughout the year.

While attending and graduating from a DAU-offered course is the most common method for people to meet the training standard associated with their position, two other options exist. The first is to complete a DAU-approved equivalent course. The second is to receive a fulfillment credit via the DAU fulfillment program. DAU-provided definitions for equivalency and fulfillment are as follows:

- **Equivalency:** DAU provides the opportunity for other organizations (colleges and universities, DoD schools, other federal agencies, commercial vendors, and professional societies) to offer courses, programs, or certifications that DAU will accept as equivalent to one or more DAU courses if, upon evaluation of the organization's materials and standards, they adequately address the DAU course learning outcomes for a select DAU course or courses.
- **Fulfillment:** The fulfillment program permits the assessment of a workforce member's demonstrated competencies (capabilities acquired through previous training, education, and/or experience) against the learning outcomes/objectives of select DAU courses.

We will discuss in the next chapter how we incorporate historical DAU course enrollment, equivalency, and fulfillment data into our forecasting model.

C. Contents of Paper

Chapter 2 describes in some detail many of the methodologies explored and those eventually used in constructing the demand forecast for acquisition training courses. This chapter also provides information on all of the data that were collected and derived to support the analyses. Chapter 3 contains the results of the forecasting model for fiscal year (FY) 2018 and FY 2019. It also shows the comparisons we constructed between the IDA-derived forecast and the previous forecasting approach.

2. Methodology

This chapter outlines the overall methodology, data sources used, data processing conducted, and some preliminary analyses conducted. It also provides an in-depth discussion of the methodology used to create the demand forecasts.

A. Methodology Overview

A multi-step process was used to create the demand forecasts, which are discussed in detail in section 2.E. First, the IDA team forecasted the demand rates for each course by personnel cohort using historical data. Second, the team forecasted the number of personnel within each cohort, partially using a machine learning (ML) model. Finally, we multiplied the forecasted cohort demand rates by the projected personnel in each cohort and summed across cohorts to predict the future demand for each course.

B. Sources Used

There were several sources of data used for this analysis. The main sources of data contained records at the individual level (non-person identifiable), supplied by the office of the DACM from in-house data systems. Several other sources were used in the course of our analyses; all are discussed in this section.

The first main source was a series of spreadsheets containing student-level data on direct interactions with the registration system, referred to as the Historical Graduates Reports. These data contained information about interactions each person had with each course in a given year, along with several individual-level fields (e.g., DAU ID, career level, career field, organization). Note that this dataset—annual data from FY 2007 to FY 2016—included course history for 4th Estate personnel only. At the conclusion of our research, we were able to update the model with FY 2017 data.

The second data source used in the analysis was a collection of annual end-of-year snapshots from FY 2005 to FY 2016 referred to as the Workforce Count Reports. These contained information about each person in the acquisition workforce within the 4th Estate for a given year, regardless of whether or not they took a course that year. Note that this dataset did not include acquisition professionals from the Services; thus, any history of 4th Estate personnel working for the Services previously is unknown. Data fields available by person and year included the following: career field, career level required, career level achieved, supervisor designator, job title, service completion date, occupational series code, intern program indicator, organization, retirement eligibility code, rule of 92, supervisor designator, age, acquisition experience, pay grade, pay plan, state, zip code, active indicator, and retirement plan code. We updated the model with FY 2017 data when we received them in October 2017.

Note the above detailed individual level data sources did not contain personnel data or course history for a select group of personnel, largely from the intelligence community. In the absence of detailed data for this community, the team was unable to model their behavior explicitly and, thus, relied on a third source of information to supplement our final estimates of course demand. This third data source was a Course Enrollment report—containing summary data from FY 2012 to FY 2016—that conveyed the total number of people showing interest (reservations, cancellations, wait list, etc.) in each course by course number, year, and organization, among others. These data were not at the detailed individual level, but did include the demand from the intelligence community.

The IDA research team also used the DAU Course Catalog, accessed on DAU's website. From this site, the team was able to gain an understanding of the courses required for each combination of career field and certification level. Additionally, we were able to determine which courses had been referred to previously using another course name, so that we could keep continuity among courses (and their level of interest over time) in our analyses. For example, BCF 107 was retired in October 2016 and renamed BCF 131; both are the Applied Cost Analysis course.

Another data source used for the final product was a count of personnel in the 4th Estate who had received credit for a course (without taking the DAU course) through either a fulfillment or an equivalency. We received data from the 4th Estate on course equivalency and fulfillments going back to at least FY 2005, matching the timeline of other data received from DACM.

C. Data Processing

The two main data files used in the analysis (Historical Graduates Reports and Workforce Counts Reports) contained individual-level records. Both sources contained corresponding fields that could be used to match records in a useful and appropriate way. This was important in merging and using the personnel data and the historical course information for each member of the workforce over time. By matching the two sources, we created a database in which there existed a single record for each unique person (DAU ID) and fiscal-year combination where they existed in the workforce. Note that, upon matching, the team found DAU ID records in the Workforce Reports that did not occur in the Historical Graduates reports, which was a signal that that person (DAU ID) did not take any courses.

After merging the workforce data and course history sources, we applied additional data processing. The substantive processes whereby significant assumptions were implemented or new data fields were created are described next.

1. Defining Demand

The first data field the team created was an overall metric of demand. Recall that we are interested in forecasting future demand for DAU training courses from 4th Estate acquisition professionals. We define demand for each course as the total number of individuals in 4th Estate agencies who would like to take a course in a particular year. Unfortunately, demand is not readily observable in the historical data; the historical data are likely limited by historical supply. For example, perhaps a particular course had 50 seats and historical data show that 50 seats were filled. That data point does not indicate whether 50 people wanted to take the course or if the demand exceeded available seats. We thus had to construct a measure of demand from the available historical data elements.

The Historical Graduates Reports provided the main source of data in constructing a measure of historical demand. This database contained an account of the following actions for each person and course:

- Input (attended the course)
- Register
- Graduate
- Wait List
- Cancel
- Attrition
- No Show
- Walk-in

For our measure of demand, we counted the number of individuals who had any of the above listed interactions with the course. Clearly, many people had more than one interaction with a course in a year. For example, a person may have registered for, attended (input), and then graduated from CON 170 in a year. The team would count this person only once as having a demand for CON 170 in that year, as we were estimating the total number of *unique* individuals, not counting all of their interactions. Thus, to each record we appended a field indicating whether or not a person showed interest in each of the DAU courses for which a forecast is provided. Incorporating counts of wait lists and walk-ins, for example, into the demand measure provides a more insightful view of the number of people who desired to take the course. However, if a person decided not to join the wait list for a fully booked course offering, this measure could underestimate the true demand.

2. Tenure

In preliminary data exploration, we found that one of the characteristics relating to whether or not a particular person was likely to take a course was the length of time the person had been in the workforce and, more specifically, in their current position. The team also noted that if a person had been promoted into that current position from elsewhere within the workforce, they may have a different likelihood of taking a course than if they were hired from outside the 4th Estate workforce. After being in a position for three or more years, the demand rate differential between those who were promoted and those who were new to the workforce was relatively small.

As a result, we added a variable within the dataset called "Tenure" that indicated both the duration of the person's time in the current position and whether the person was promoted to that position from within the 4th Estate workforce or completely new to the 4th Estate workforce. Thus, each unique person-year combination was assigned exactly one of the following tenure values:

- New – First Year
- New – Second Year
- New – Third Year
- Promoted – First Year
- Promoted – Second Year
- Promoted – Third Year
- More than three years

Note that we only examined position changes between fiscal years as of the last day of each fiscal year; thus, multiple position changes within a year were not regarded. A person was considered to be in the same position if they were in the same career field, had the same career level required, and were at the same organization within the 4th Estate.

3. Other Processing

As previously mentioned, the team needed to combine some historical course data due to the fact that course names changed over time (yet were functionally the same course). In general, we simply assigned the data contained under the previous course name with those under the current course's name. Note that historical course data for courses with a new name were listed with the new name. For this course data merge, we assumed that in the rare case where a person-year had expressed interest in both the existing course and the predecessor course, the "interest" overall would be counted just once per year.

Additionally, we filtered records to include only active 4th Estate personnel because they constitute the population in which our sponsors are interested. Inactive personnel do not meet the threshold of acquisition work required to be considered part of the active acquisition workforce.

D. Preliminary Analyses

Before developing the forecasting methodology, the team explored the data to gain situational awareness. We wanted to understand answers to questions such as: Which portions of the population are driving demand? Which courses were in most demand? How has demand changed over time? Answers to such questions would eventually help form the overall forecasting methodology.

We provided demand forecasts for nearly 70 courses. We first examined the demand (interest) for these courses over time, which is displayed in Figure 2, parsed by course (legend not shown).

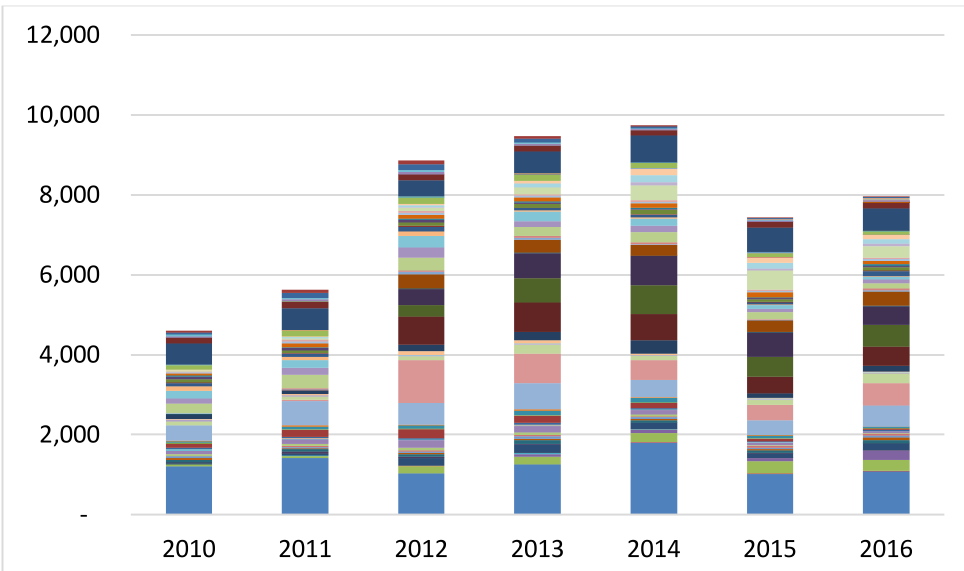

Figure 2. Overall Demand Parsed by Individual Courses over Time

An increase in demand can be seen from the beginning of our dataset until a peak demand of almost 10,000 seats (for these select courses) in FY 2014, followed by a slight decrease in the years FY 2015 and FY 2016. New course requirements, particularly in the contracting field at the start of FY 2012, drove some of the increased demand in peak years.

Overall, the demand was spread across numerous courses. The course with the largest demand is ACQ 203 (shown in blue at the very bottom of the chart), with an annual demand of just over 1000 seats, while very small courses might enroll less than 10 people per year. It is evident that demand was coming from many different courses. Upon closer examination of individual courses, we found that some courses had increasing demand over this time period at the same time that others were decreasing.

After a check of overall demand, the team began to take a closer look at the possible influences of overall course demand. First, we examined the workforce over time. Figure 3 shows the active 4th Estate workforce over the last seven years of data by career field.

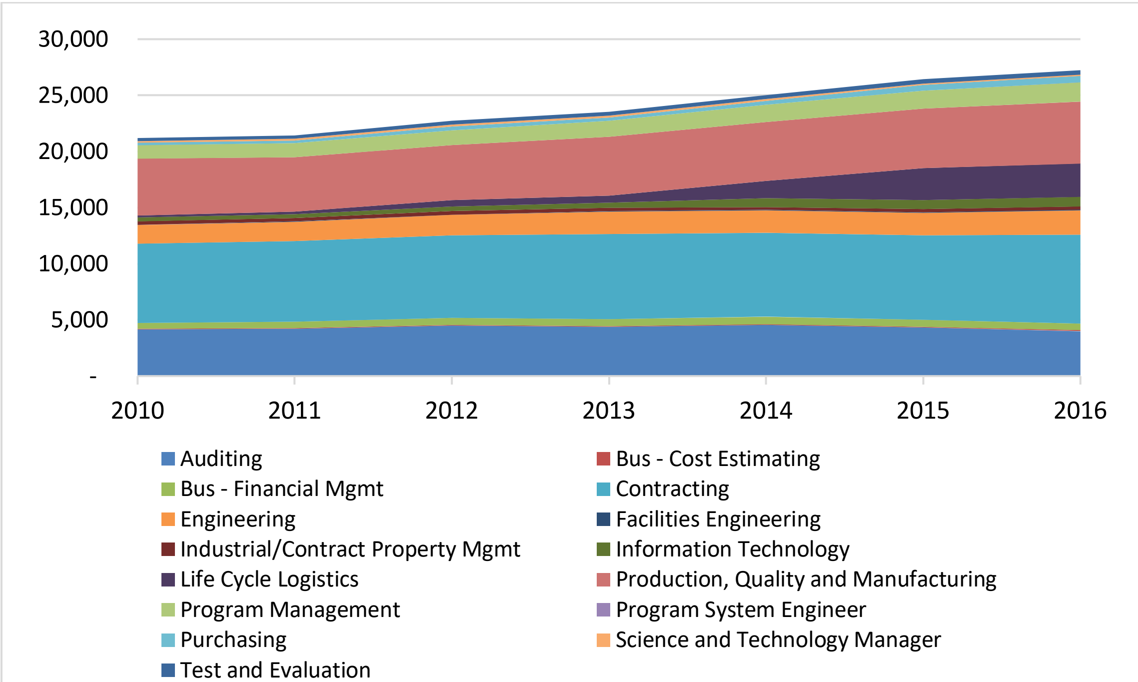

Figure 3. Active 4th Estate Workforce, Shown by an Estimate of Years in Position

The workforce has steadily increased from FY 2010 (just over 20,000 people) through FY 2016 (just over 27,000 people). There has been notable growth in the Life Cycle Logistics career field from FY 2013 (just over 600 people) to FY 2016 (almost 3000 people). Most other career fields experienced relatively small growth or held a constant workforce size.

Figure 4 shows the number of active personnel in the 4th Estate by the career level required for the person's current position. Note this does not necessarily mean the certification level achieved, but what the person was required to achieve within 24 months of starting their position. The workforce was largely made up of Level 2 and 3 positions, with Level 1 positions making up only about 6 percent of positions over this time period.

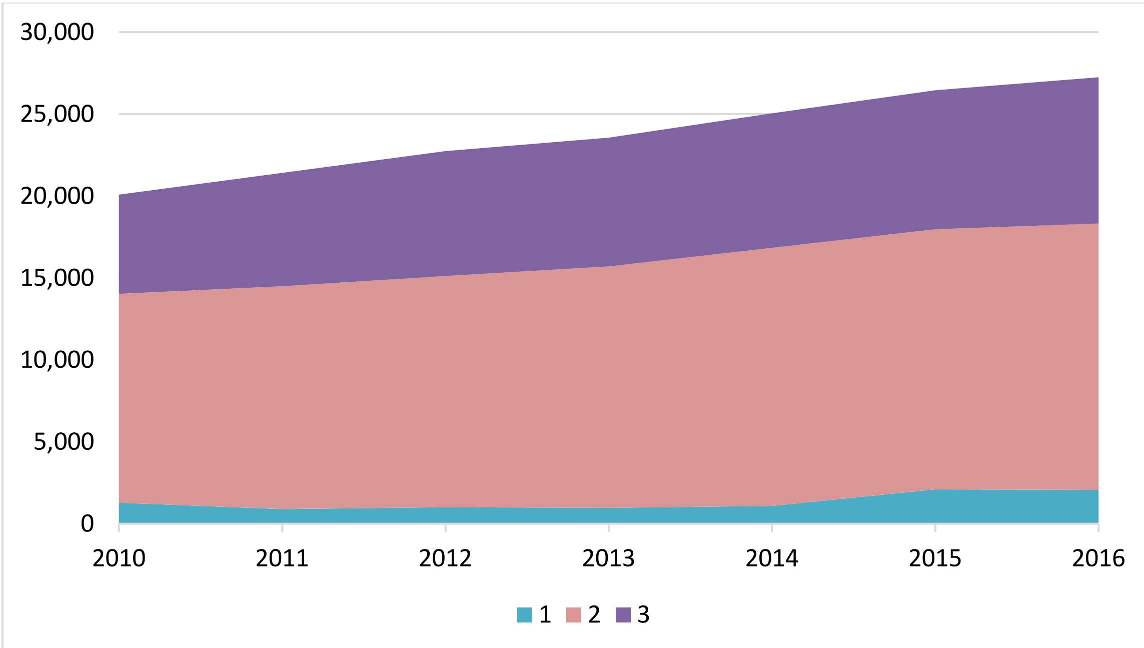

Figure 4. Active 4th Estate Workforce, Shown by Career Level Required

Since personnel new to their positions are required to achieve certification within 24 months of beginning their position, we were also interested in the length of time the workforce members had been in their position. Presumably, those in their positions fewer than two years would be likely to take courses. Figure 5 shows the number of active personnel in the 4th Estate parsed by years in position.

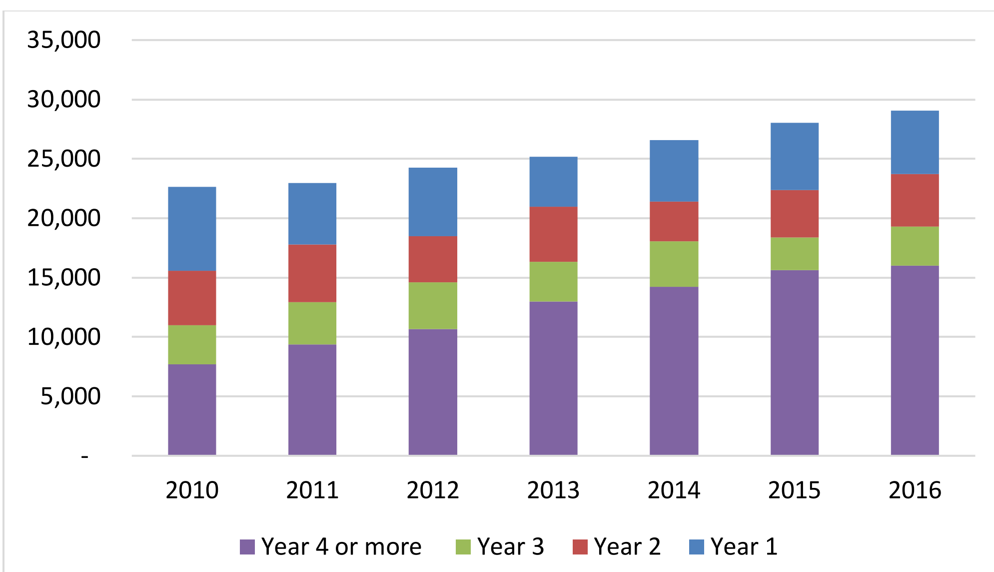

Figure 5. Quantity of People in the Active 4th Estate Workforce, Shown by Years in Position

A large, but decreasing, portion of the population appears to be in their position three or fewer years—roughly 65 percent in FY 2010 and steadily decreasing to 45 percent in FY 2016. The portion of personnel who are within their first 24 months of starting a new position was about 50 percent in FY 2010 and decreased to about 33 percent in FY 2016.

To explore whether the number of years in position was driving demand, we next examined the actual historical demand by the years in position, found in Figure 6.

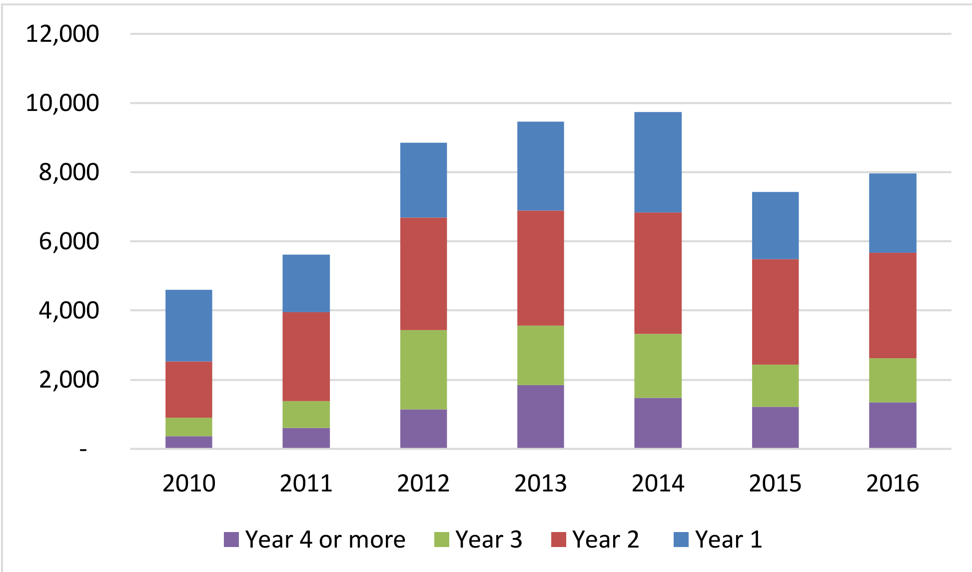

Figure 6. Course Demand by Years in Position of the Active 4th Estate

As the figure demonstrates, the majority of the demand is driven by those in their positions for fewer than three years (comprising roughly 85 percent during the time period shown), and, moreover, is driven to a large extent by those in their position for less than two years (about 67 percent). This was important as the team set up our forecasting methodology.

Next, we hypothesized that personnel who were promoted to their current positions from within the 4th Estate workforce might be somewhat less likely to take courses than those coming from outside the 4th Estate workforce. The reasoning behind this hypothesis was that those personnel who had already been in the workforce were likely to already hold a certification (or at least have started satisfying one), perhaps even within their current field, and would therefore not require as many pre-requisite courses or additional training as someone who was joining from outside the workforce. To corroborate this, team members compared the overall demand separated into these two groups. Data showed differences in demand rates between these two groups over time for many courses.

From this exploratory analysis, the research team learned some important facts about our problem and the available data. We learned that demand has been increasing and is spread across many courses. We found that the duration of time a person has been in their position and whether or not they came from inside the workforce are two important factors in estimating future demand.

E. Forecasting Approach

1. Challenges to Forecasting

One of our forecasting challenges was to create a course demand forecast for year *t+2* where the most recent complete fiscal year of data available was for year *t*. This is due to the time horizon in which the 4th Estate must create their forecasts.

Another challenge of the forecasting methodology and a requirement of the model was that it should not only calculate the most likely demand forecast, but should also have the ability to forecast demand under alternative input assumptions, effectively answering "what-if" scenarios for the model user. Thus, the model needed to (1) receive alternative model inputs and (2) re-compute an overall demand forecast based on the new input. For example, the DACM indicated that the size of the workforce (by agency, career field/level) would be one important input variable, as workforce size often increases or decreases from year to year.

As discussed, the 4th Estate provided the IDA research team with a wealth of variables for each record (person-year) in the database. The team speculated that such rich data would lend themselves well to an ML solution. As such, we attempted several model structures to predict demand. One initial test of an ML solution used the workforce in year *t* to predict the course demand for year *t*. While that model was rather successful and demonstrated that ML could accurately predict course demand, it did not satisfy the constraints of the problem; it used two years of workforce data that would not be available in the realistic time constraints under which the 4th Estate must create its forecasts.

A key conclusion from our initial analysis is that traditional ML algorithms are not the best option for predicting 4th Estate course demand. Recall from Figure 4 that roughly two-thirds of the course demand is generated by personnel who have been in the workforce for less than two years. In addition, the most recent course history and workforce data available for predictions of year *t* were data from year *t-2*. The combination of these two facts implies that detailed information about the people supplying a majority of the demand is unknown, as they have not yet moved into their position. Therefore, insufficient data are available for applying such an ML model to directly predict course demand.

2. Final Methodology

To satisfy the aforementioned challenges, the IDA team developed a methodology that separates the problem into two distinct parts. Summarized simply, the two significant portions of the problem are (1) forecasting the demand rate for each of the 70 courses and (2) forecasting the workforce. Each of these forecasting problems is done at a population subgroup, or cohort, level. These two values are then multiplied to yield the overall demand forecast for each course. The following sections will discuss the methodology in further detail.

a. Cohort Selection

As previously mentioned, we forecasted the workforce and the demand rates for subgroups (or cohorts) of the population, rather than the population as a whole. This enabled us to tailor course demand rate forecasts to each subgroup as appropriate, which allows for a more precise overall estimate. In particular, it also allows for a more accurate forecast within the delivered software tool with which a user will evaluate various assumptions.

We subdivided the workforce into cohorts according to the characteristics that influence the rate at which courses are taken. For example, as discussed previously, the team discovered that the number of years a person has been in their position influences their likelihood to take a course (i.e., employees in positions for four years or more were less likely to take courses generally, while those in their first or second year were likely to take courses). Another example characteristic is whether an individual was promoted to their current position of employment from elsewhere within the 4th Estate acquisition workforce or had not been part of the 4th Estate workforce previously. The overall demand rate was driven by these characteristics; thus, they were part of the cohort groupings.

In addition to "years in position" and "promoted/new," variables used to define the cohorts included career field, career level required, and agency. Thus, there were five variables defining the cohorts, shown in Figure 7, along with relevant subscripts used in equations that follow.

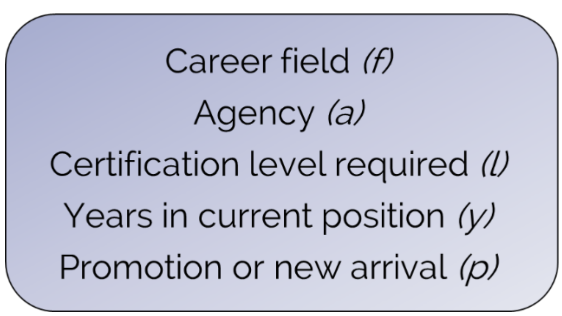

Figure 7. Variables Defining the Cohorts

The following two sections discuss the methodology we used to solve the two distinct forecasting problems: forecasting demand rates and forecasting workforce.

b. Forecasting Demand Rate

In order to forecast the demand rate, we first summed the total demand by year and parsed that by cohort groups. We then summed the historical workforce by cohorts per year. Next, we calculated a three-year average for the demand rate for each cohort by dividing the total demand for the three years by the total workforce for the three years. Note that if the course was not offered in a particular year, we zeroed out that year's workforce in the denominator so that the demand rate would not be incorrectly reduced. Equation (1) shows the formulation of this calculation, where the subscript *j* denotes the course, while the other subscripts correspond to those found in Figure 7. Note that the workforce in the denominator is actual workforce in historical data, not the projected workforce.

$$DemandRate_{falypj} = \frac{\sum_{t=1}^{n} Demand_{falypjt}}{\sum_{t=1}^{n} Workforce_{falypt}} \tag{1}$$

While summing the total demand and total workforce to be used in the demand rate calculations, we did encounter some issues with small subgroups. Most cohorts driving demand are rather large. Some cohorts, however, can get quite small—perhaps having fewer than 10 people. This becomes a problem when calculating historical demand rates for future application. For example, imagine a cohort has only two people in it and both of those people happened to take a particular course. The demand rate (for this single year of data) would be 100 percent. If we are applying historical demand rates to estimate future course demand, is it reasonable to assume that 100 percent of this cohort will take the course in the future? We think not. To overcome this issue we aggregated some cohorts together to form larger buckets with which to calculate demand rates. That is, we aggregated some cohorts by agency such that the subscript *a* in the demand rate equation was represented by a group of agencies (called "Other") rather than individual agencies. This allowed for a more robust and accurate estimated demand rate.

This method inherently allows the historical data to determine the demand rate, rather than comparing DAWIA requirements per certification field and career level required against what a particular person has already taken. The methodology captures demand from personnel who take courses for reasons other than certification for their current position, such as prospective career advancement or dual certification. We assume that a particular cohort's propensity to take a course in the future is equal to what that same cohort's propensity has been over the last three years.

c. Forecasting Future Workforce

Next, we used ML to help forecast the future workforce to which the demand rates would be applied. More specifically, in the first step of forecasting the overall future workforce, the team used ML to predict whether or not each person in year t would be in the same position in year $t+2$. Note that the dependent variable being predicting here (essentially attrition) is different from predicting course demand directly in the previously discussed ML model, which violated problem assumptions. This problem uses existing data to predict the attrition two years ahead.

To do this, the team set up a dataset with which to train the model. We added a 0/1 variable to the historical data, which indicates if that person is in the same position in two years. This became our dependent variable for the ML model. The independent variables are all other known variables from the workforce dataset. A depiction of the problem formulation can be found in Figure 8.

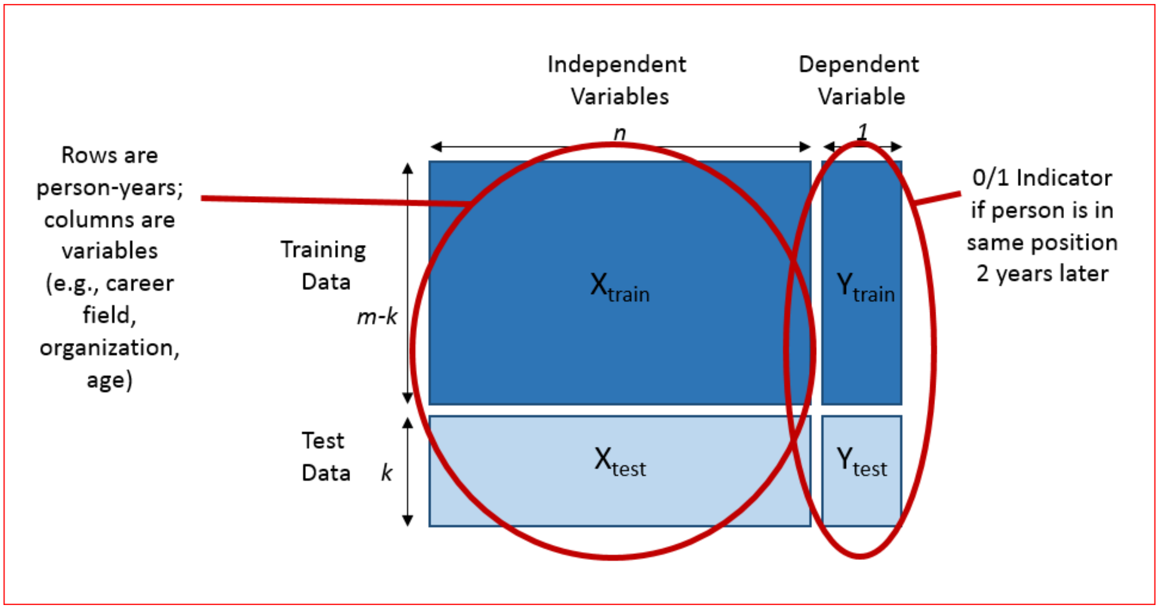

Figure 8. Problem Set-up for Building the ML Model to Predict Whether or Not a Person Remained in Position Two Years Later

We gathered several years of the most recent data available for which we were able to construct the dependent variable. We then trained a gradient boosting machine (GBM) algorithm with the training set. This algorithm is a decision tree-based method that ensembles many weak learning models together to form a more predictive model. It is a robust methodology with several advantages: it generally performs well on a variety of problem types, it approximates interactions and non-linear transformations, and automatically handles missing values. It is also able to implicitly handle correlated independent variables (e.g., age and Rule of 92).

Once the model was constructed, the team was able to assess the model results by applying it to test sets of data that are out of sample and a different time period from the training dataset. We experimented with various assumptions of the number of years of historical data included in the training dataset. We compared the model predictions with the actual results and found that three years of historical data provided the best results. We then removed any variables from the dataset that were not contributing to the model and re-ran the GBM algorithm to create the final model. The relative influence of the top variables in the final model is shown in Figure 9.

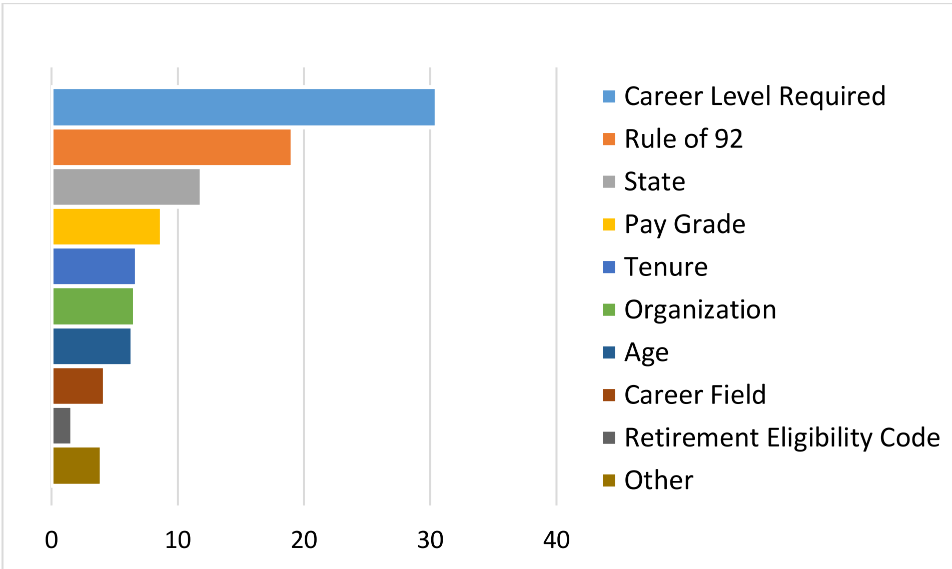

Figure 9. Relative Influence of Variables in Final Model

The variables selected by the model appeared to be reasonable and logical in predicting whether a person is likely to leave their current position. Career level required was likely influential, particularly for the workforce in Level 1 (who are likely to move on to Level 2 within two years) and workforce in Level 3 (who are likely to remain in these more senior positions). Rule of 92 is a variable combining both age and years of experience that partially determines eligibility for retirement. Age and years of experience are likely good predictors of retirement but also could help predict which employees are likely to be promoted from Level 1 positions (those in lower age/experience brackets). Thus, it was unsurprising that Rule of 92 would be a strong indicator in this particular model. Note that although the variables Rule of 92, age, and retirement eligibility code are likely somewhat collinear, the GBM methodology implicitly accounts for such collinearity in creating its individual decision trees and the overall model.

The team then applied this model trained on historical data to the independent variables in the dataset for year *t* to get an estimate of the people who will remain in year *t+2*. Those who will be in the same position were counted in the projections for year *t+2* and their tenure status was updated.

Once we estimated the number of people who depart their positions, we assumed a one-for-one replacement by position (career field, career level required, and agency). This assumption inherently implies that the total number of people in the workforce remains the same as in the base year to which the ML model is applied. That is, if the ML model is applied to base year FY 2016 in order to forecast the workforce in year FY 2018, the FY 2018 forecast will automatically have the same number of people (and, furthermore, the same number of positions) as the year FY 2016. Note that the software tool user can make changes to this assumption by increasing the level of personnel across the entire workforce or for particular cohorts.

Next, the team assigned the replacement personnel to a tenure category based on historical data. That is, for the three most recent years of data, we calculated the distribution across tenure groups among personnel who were in the first two years of their position. We applied this distribution to the replacement personnel and added them to the people who remained in position from year *t* to year *t+2*. We then had our final projection of the workforce by individual cohort.

d. Forecasting Total Course Demand

Next, we combined the workforce and demand rate calculations to estimate the total demand for each course by year. The formula for the total demand *D* by course *j* in year *t* is given by:

$$D_{jt} = \sum\left[\boldsymbol{W}_{falypt} \times \text{DemandRate}_{falypj}\right] + \mathrm{I}_{jt} + \mathrm{E}_{jt} \tag{2}$$

The calculation for each cohort is summed across all cohorts for each class in each forecasted year.

Once the demand calculations for cohorts were complete, the team then added two other components of course demand. The first was the demand from the intelligence community that we calculated separately by year and course as appropriate. We assumed that the intelligence demand for the forecasts would be equivalent to the average of the five most recent years of available actual data. The demand from intelligence professionals was quite stable by course over the time period FY 2012 to FY 2016.

Second, we added in the counts of data from the equivalency courses; we assume that equivalencies were used when course capacity at DAU was insufficient, causing a person to satisfy the course requirement elsewhere. We assumed the equivalency demand would be equal to the average of the five most recent years of data. Equivalencies showed more variability than the intelligence demand, likely associated with the available capacity by year and course; however, the overall demand from equivalencies was small enough that it did not warrant a separate methodology to attempt to predict more accurately. Note that we did not consider fulfillments to be additional demand, since those people would not likely have taken the fulfilled course regardless of whether there was an available seat.

Once each of these portions of the estimate was complete, they were summed together to yield the overall forecast. The next chapter shows reviews the results of our forecasting methodology and shows the forecasted course demand for FY 2018 and FY 2019.

3. Results

A. IDA Forecast

Although the model deliverable is capable of conducting sensitivity analyses based on user input, the IDA team created a baseline forecast that will serve as the model's starting position, prior to any input alternatives by the user. This baseline forecast is IDA's best estimate of course demand in the future and assumes no change in the total number of people in the workforce from the last year of available data. Thus, the forecast for FY 2018 assumes the same overall workforce as that in FY 2016 and the forecast for FY 2019 assumes the same overall workforce as that in FY 2017. The demand rate calculations for the FY 2018 forecast use historical data from FY 2014 to FY 2016 while the forecast for FY 2019 uses historical data from FY 2015 to FY 2017.

Figure 10 shows total demand (parsed by course) for the years FY 2012 to FY 2019, where FY 2018 and FY 2019 are IDA forecasts and the previous years are actual results. Table 1 shows the resulting total forecast by course number for forecasted years FY 2018 and FY 2019.

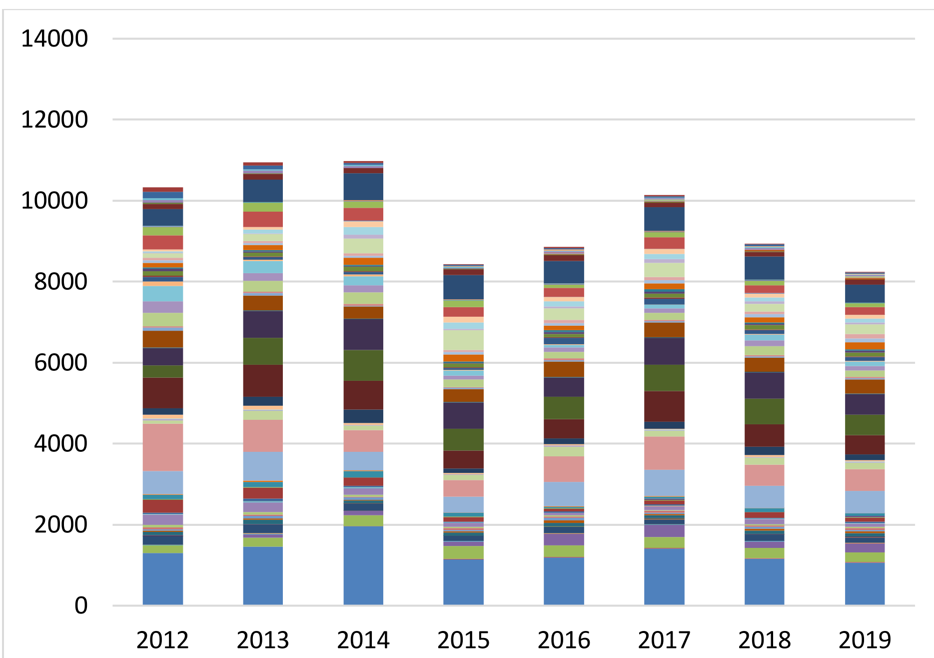

Figure 10. Actual Course Demand for the Years FY 2012 through FY 2017 and Forecasted Demand for Years FY 2018 and FY 2019

Table 1. Actual Course Demand for FY 2017 and Forecasted Course Demand for Years FY 2018 and FY 2019

Course	2017	FY 2018	FY 2019
ACQ-203	1233	1155	1057
ACQ-230	12	12	14
ACQ-265	240	259	250
ACQ-315	283	165	215
ACQ-340	8	8	7
ACQ-350	2	3	3
ACQ-370	112	156	133
ACQ-380	17	0	17
ACQ-404	7	2	5
ACQ-405	15	11	11
ACQ-450	62	79	78
ACQ-451	27	44	39
ACQ-452	28	54	45
ACQ-453	42	38	41
BCF-131	19	31	26
BCF-205	38	115	92
BCF-206	12	14	11
BCF-209	4	4	3
BCF-215	28	22	21
BCF-225	39	137	106
BCF-230	15	8	7
BCF-250	30	-	29
BCF-301	29	86	63
BCF-330	5	9	4
CON-090	605	542	550
CON-170	773	532	543
CON-232	135	164	158
CON-234	17	18	18
CON-243	11	8	8
CON-244	19	42	39
CON-252	167	204	133
CON-270	721	553	488
CON-280	619	627	500
CON-290	627	652	510
CON-334	8	13	15
CON-360	323	354	340

Course	2017	FY 2018	FY 2019
CON-370	39	28	33
COR-222	23	35	27
ENG-202	146	218	161
ENG-301	100	145	122
EVM-202	45	150	105
EVM-262	3	17	10
EVM-263	132	88	103
FE-302	21	15	17
GRT-201	90	105	94
IND-105	53	34	39
IND-205	54	31	33
ISA-201	101	138	176
ISA-301	63	68	106
ISA-320	56	72	104
LOG-201	357	192	239
LOG-211	81	49	52
LOG-340	136	108	95
LOG-350	114	91	88
LOG-465	5	9	5
PMT-257	181	187	182
PMT-360	84	117	90
PMT-400	13	14	14
PMT-401	13	13	12
PMT-402	2	3	2
PQM-201B	586	565	433
PQM-301	111	123	124
RQM-310	4	11	12
RQM-403	0	1	0
RQM-413	-	-	-
SBP-102	25	25	24
SBP-202	12	9	9
SBP-210	5	16	11
SBP-301	25	-	25
STM-203	17	35	30
STM-304	9	28	26
TLR-350	5	18	14
TST-204	37	37	32
TST-303	24	26	20

Some of the courses show zeros for their forecasted values. This is most likely because the demand rates for the years on which the estimate is based are also zero. These are generally new courses that have little to no historical demand in the model's input data. As more years of demand data are added to the model, future forecasts will increase from zero in response.

B. Comparison to Other Methods

This section compares the IDA forecasting results to forecasts from other sources. This allows us to assess whether the model provides any improvement in forecast accuracy. Note that the IDA forecast for FY 2018 uses the data available through FY 2016, but not FY 2017 data, as they would not be available at the time the 4th Estate is required to complete their initial course demand forecasts. Similarly, the forecast for FY 2019 uses all known data through FY 2017, but nothing after.

One of the forecasts to which we compare the IDA results is the 4th Estate forecast, which is completed almost two years prior. Figure 11 shows the results of the IDA methodology and the 4th Estate forecasts compared to the actual course demand for two years: FY 2015 and FY 2016.

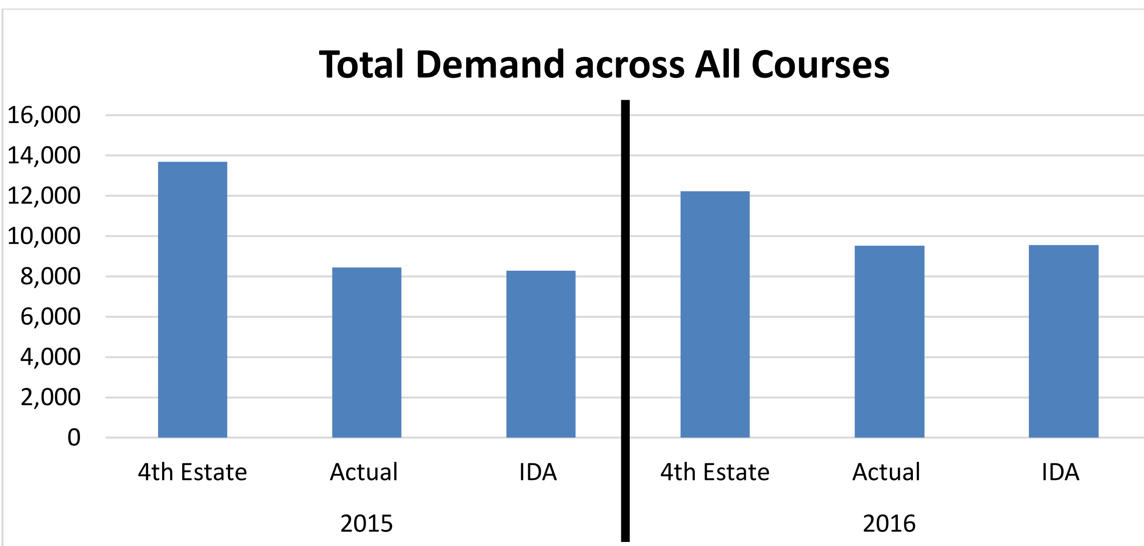

Figure 11. Course Demand as Forecasted by IDA and the 4th Estate as well as Actual Course Demand for Years FY 2015 and FY 2016

In addition to comparing IDA results to the 4th Estate forecast, the team also compared results to two other estimates. One point of comparison was simply the last actual data point for course demand. Another point of comparison was a weighted average of the two most recent years of historical data, with two-thirds of the weight going to the most recent year and one-third of the weighting going to the data from the year prior to that.

If we examine the difference between each of these four forecasts and the actual course demand, we can gain a sense of how well each of the methodologies estimates the demand. We did not compare demand for courses where the 4th Estate did not provide a forecast. Figure 12 shows the details of the metric absolute difference for the comparisons conducted (and shown in Figure 13).

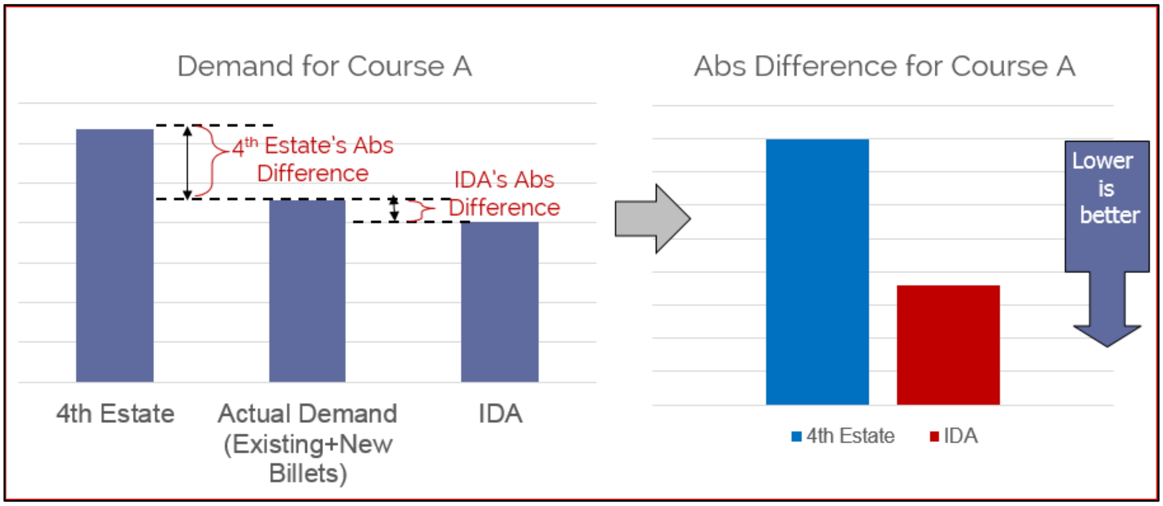

Figure 12. Demonstration of the Metric Absolute Difference, used for Comparing Various Methods

Figure 13 shows how close each of the four forecasts are to the actual value using absolute difference for 30 of the most demanded courses. For these comparisons we have combined forecasts for years FY 2015 and FY 2016. Note that a lower number demonstrates closer to the actual and a better forecast; a perfect forecast would show a value of 0 in Figure 13.

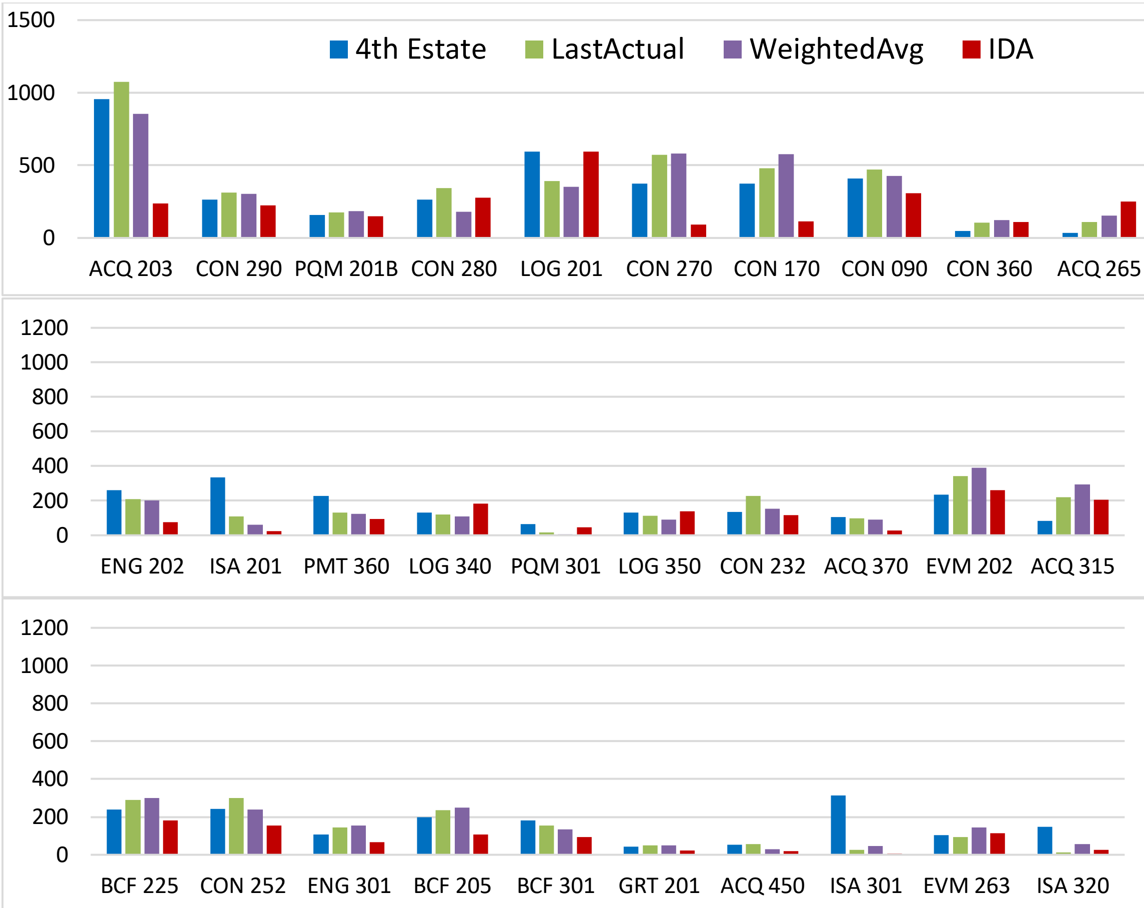

Figure 13. Comparison of Various Course Demand Forecast Estimates to the Actual Course Demand for 30 of the Most Demanded Courses

As can be seen in Figure 13, the IDA forecast is closer to the actual demand than the 4th Estate forecast as well as the latest actual and weighted average for many of the courses (shows a lower value), notably for ACQ 203, CON 270, and CON 170, which are some of the most popular courses. Some slightly less popular courses, but where the IDA forecast still shows an improvement over the 4th Estate forecast and other forecasting methodologies, include ISA 301, ISA 320, BCF 205, ACQ 370, ENG 202, and PMT 360. The IDA methodology tends to outperform the latest actual and weighted average forecasts as well. Some of the courses for which the IDA methodology did not forecast as well as the 4th Estate in FY 2015 and FY 2016 include three logistics courses: LOG 201, LOG 350, and LOG 340. We found that the IDA methodology largely underestimated the actual course demand for these logistics courses. After investigating further, the team found that the underestimation was largely driven by a very recent increase in personnel in the workforce in the Life Cycle Logistics career field that was not visible to our model due to the forecasting horizon. We were able to parse the portion of the actual demand for those courses due to the increased workforce and found a significant portion of the difference to be due to the additional workforce. A similar situation occurred for some courses in the acquisition field (ACQ 203 and ACQ 265), where a recent increase in personnel caused the IDA forecast to underestimate the total demand.

Figure 14 shows another way to view the IDA results in comparison to the current 4th Estate methodology. The figure shows a histogram of the difference between the actual and forecasted values for both the IDA methodology and the 4th Estate's methodology. Data points counted in the histogram are the difference by course for each of the years from FY 2015 to FY 2017. IDA's forecasts tend to be centered near 0 (no difference from actual) while the 4th Estate forecasts tend to be biased high (higher than the actual).

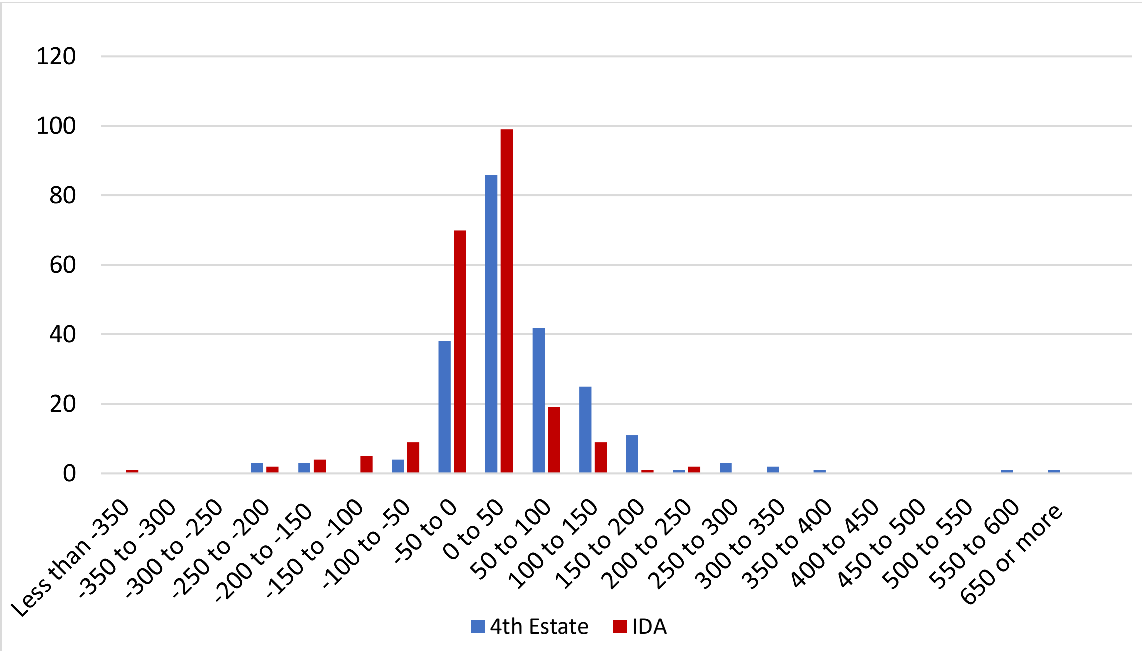

Figure 14. Histogram of the Differences between the Forecasts and the Actual Demand by Course across the Years FY 2015 to FY 2017

C. Summary

IDA developed a methodology for forecasting demand for DAU training courses from DoD 4th Estate acquisition professionals. IDA's forecasting methodology demonstrated a pronounced improvement over the old forecasting approach for the years FY 2015 and FY 2016. Overall, the IDA forecast was about 49 percent closer to the actual course demand than that of the 4th Estate. For the courses for which the IDA forecast was better, it averaged about 100 seats (42 percent of class size) closer to the actual demand.

IDA also developed an interactive computer model. The user-friendly interface developed for this model allows for easy viewing of the forecast while also enabling users to easily perform "what-if" analyses. This capability provides the user the ability to see the likely implication of changes to the size and composition of the workforce as well as changes to demand rates due to events such as hiring freezes or travel bans.

With a more accurate course demand forecast, the 4th Estate DACM is better positioned to ensure access to desired training courses for the 4th Estate acquisition workforce.

Illustrations

Figures

Table

References

Course Enrollment Database, FY 2012 to FY 2016. Received from Jonathan Higgins, Office of the Director Acquisition Career Management, March 2017.

Fulfillments and Equivalency Data, FY 2005 to FY 2016. Received from Jonathan Higgins, Office of the Director Acquisition Career Management, October 2017.

GAO. "Defense Acquisition Workforce: Actions Needed to Guide Planning Efforts and Improve Workforce Capability." Washington, DC: GAO, December 2015.

Historical Grads Report by Person, FY 2007 to FY 2016. Received from Jonathan Higgins, Office of the Director Acquisition Career Management, March 2017.

Historical Grads Report by Person, FY 2017. Received from Jonathan Higgins, Office of the Director Acquisition Career Management, October 2017.

Nataraj, Shanthi, Christopher Guo, Philip Hall-Partyka, Susan M. Gates, and Douglas Yeung. *Options for Department of Defense Total Workforce Supply and Demand Analysis: Potential Approaches and Available Data Sources*. Santa Monica, CA: The RAND Corporation, 2014.

Workforce Count Report, FY 2005 to FY 2016. Received from Jonathan Higgins, Office of the Director Acquisition Career Management, March 2017.

Workforce Count Report, FY 2017. Received from Jonathan Higgins, Office of the Director Acquisition Career Management, October 2017.

Abbreviations

DACM	Director, Acquisition Career Management
DAU	Defense Acquisition University
DAWIA	Defense Acquisition Workforce Improvement Act
DoD	Department of Defense
FY	Fiscal Year
GBM	Gradient Boosting Machine
IDA	Institute for Defense Analyses
ML	Machine Learning

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
xx-04-2018	Final	Jul 2016 - Nov 2017

4. TITLE AND SUBTITLE
Forecasting DAU Course Demand for the 4th Estate

5a. CONTRACT NUMBER
HQ0034-14-D-0001

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Hunter, David, E.
Breuer, J., M.
Giles, Margaret, L.
Guerrera, Kristen, M.
Kane, John, J.
Oh, EunRae
Rieksts, Brian, Q.
Tate, David, M.
Wake, B., Douglas
Wu, Kevin, Y.

5d. PROJECT NUMBER

5e. TASK NUMBER
FJ-7-4212

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, VA 22311-1882

8. PERFORMING ORGANIZATION REPORT NUMBER
IDA Paper P-8788
Log: H 17-000589

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
4th Estate (Acting) Director Acquisition Career Management
9820 Belvoir Road
Building 205, room 117
Fort Belvoir, VA 22060

10. SPONSOR/MONITOR'S ACRONYM(S)
DACM

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The objective of this paper is to forecast Defense Acquisition University (DAU) course demand by civilian acquisition professionals in 4th Estate agencies. The forecasting model uses a two-stage approach to forecast demand. The first stage forecasts the workforce levels by cohort by first estimating attrition and promotions in the workforce using a machine learning algorithm called a gradient boosting machine. We then assume a one-for-one replacement of positions vacated using historical workforce data. The second stage forecasts the rates at which demand occurs by cohorts of the workforce using historical demand rates for the last three years available at the time of the forecast and assuming those average cohort demand rates hold for the future. We define workforce cohorts by combinations of career field, certification level required, agency, years of experience, and previous status (new hire/promotion). The forecasted total demand for each course is calculated by multiplying the workforce levels by the forecasted demand for the course by cohort. Benchmarking against 2015 and 2016 history, this methodology, developed by the Institute for Defense Analyses, increases forecasting accuracy by nearly 50 percent. For particular courses, we saw an improvement over the historical DAU methodology in 50 of the 65 courses. We embedded this methodology in a software tool that summarizes historical data, generates forecasts for future demand, and gives the user the capability to adjust the forecast based on user-defined input.

15. SUBJECT TERMS
personnel; predictive analytics; DAU; machine learning; forecasting; gradient boosting machine, GBM

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	Same as Report	37	Bauer, Scott

19b. TELEPHONE NUMBER *(Include area code)*
(703) 806-4470

Reset