

# Analysing and Refining Pilot Training

Bruno Emond

National Research Council Canada  
1200 Montreal Road, Ottawa,  
ON, Canada. K1A 0R6  
1-613-993-0154

bruno.emond@nrc-cnrc.gc.ca

Cyril Goutte

National Research Council Canada  
1200 Montreal Road, Ottawa,  
ON, Canada. K1A 0R6  
1-613-993-0805

cyril.goutte@nrc-cnrc.gc.ca

Scott Buffett

National Research Council Canada  
46 Dineen Drive, Fredericton,  
NB, Canada. E3B 9W4  
1-506-444-0386

scott.buffett@nrc-cnrc.gc.ca

Ruibiao Jaff Guo

CAE Defense & Security  
1135 Innovation Dr, Kanata,  
ON, Canada. K2K 3G7  
1-613-247-0342

jaff.guo@cae.com

## ABSTRACT

Competency based training has become a major thrust in the development of instruction in both civilian and military pilot training. This paper reports on a joint effort by CAE and the National Research Council to identify data analytics methods relevant for the analysis, and refinements of competency based pilot training. In particular, these methods aim to identify correlations between 1) student actions and behaviours while engaging in training, and 2) students' success and incremental progression in the corresponding competencies being acquired. The paper presents some of our main results in applying sequence mining and additive factor modelling to small sets of pilot training data.

## Keywords

Aviation pilots, competency-based training, sequence mining, additive factor models.

## 1. INTRODUCTION

Over the years, CAE has developed many research collaborations with universities and government research laboratories. The current paper presents some results from a project between CAE<sup>1</sup>, the Advanced Technologies for Learning in Authentic Settings (ATLAS) research team from McGill University, and the Learning and Performance System Support program at the National Research Council Canada. The research efforts were focused on the identification of education data mining methods with practical outcomes for the improvement of pilot training. The main objective is to be able to analyse performance, and use competency models in order to refine simulation scenarios and CBT courseware. The contributions to the project represent different perspectives from sequence mining (descriptive method), to logistic regression models (predictive method). The objective was to explore the data from different points of view.

The following section presents an overview of the main trends in pilot training including competency, evidence, and scenario-based training. The next section briefly presents the data set that was used for all the analysis, and the remaining two sections presents

the main results of applying sequence mining and additive factor modeling to this data.

## 2. TRENDS IN PILOT TRAINING

To address the challenges of pilot training in the early 2000s, civil aviation stakeholders like the Civil Aviation Safety Alert (CASA), the International Civil Aviation Organization (ICAO), and concurrently the United States Air Force (USAF) have been promoting competency and evidence based training as a training model [1]–[3]. This position was in reaction to hours-based training where the number of flight hours or sorties done by a pilot determined flight or mission readiness. With the increase of flight operation complexities, it became obvious that achievement of a certain performance level on a task would be a better indication of a pilot competency, than the number of hours of practice, even though flight hours could be an indirect measure of a competency level.

There are many views about what a competency is. The International Civil Aviation Organization defines a competency as “a combination of skills, knowledge and attitudes required to perform a task to the prescribed standard” [4]. The USAF has developed an elaborate competency framework [5]. The Mission Essential Competencies (MEC) framework is intended to blend training task lists, and mission essential task lists. The MECs incorporate a wide range of pilot competencies, beyond the operational requirements, to include teams and inter-team competencies [3]. The Federal Aviation Administration (FAA) also recognizes that pilot competencies need to be defined at a higher-level than simply the low-level operations of an aircraft, especially with the increased level of automation because automated systems are not adapted to unforeseen situations [6]. Competency frameworks are usually the result of an analysis performed by subject matter experts who identify key competencies based on standards of performance and means to measure them.

Another important trend in pilot training is evidence-based training. The ICAO defines evidence-based training as “Training and assessment based on operational data that is characterized by developing and assessing the overall capability of a trainee across a range of core competencies rather than by measuring the performance in individual events or manoeuvres” [1]. The essential element evidence-based training introduces to competency based-training is the reference to operational data as a means to identify key competencies, in addition to the analysis

<sup>1</sup> <http://www.cae.com/about-cae/corporate-information/faq/>

performed by subject matter experts. Evidence-based training applies the principles of competency-based training for safe, effective and efficient airline operations, while addressing safety threats. The term evidence refers to the fact that safety threats are identified from actual flight monitoring data, such as those provided by the Flight Operational Quality Assurance (FOQA) program, Aviation Safety Action Program (ASAP) data for business aviation [7], as well as Automatic Dependent Surveillance-Broadcast (ADS-B) data.

A literature review also revealed that a combination of competency, evidence, and scenario-based training approaches can form the basis for the next generation of pilot training system. The combination requires links between the development of simulated scenario events and performance measures, both driven by training objectives [8]. This combination is well integrated in the specification of evidence-based training as defined by the ICAO [1], and the focus on scenarios and simulations provides the foundation of a strong learner centred approach.

Simulation scenarios are central to evidence-based training as the main instructional content a trainee pilot interacts with, for evaluation and learning. The approach is consistent with the principles of situated learning theory, which argues that learning best takes place in the context in which it is going to be used. Scenario-based training is mostly suitable for procedure-oriented tasks requiring decision-making and critical thinking in complex situations, and is learner centered as the scenario provides a unique opportunity for the trainee to perform and acquire competencies based on his/her competency level.

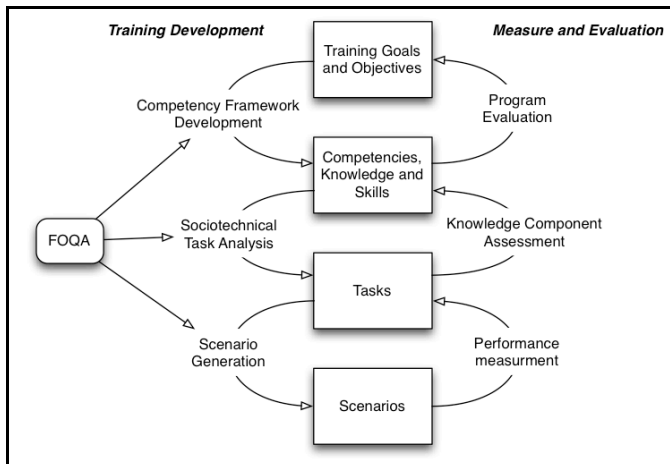


Figure 1. Competency, evidence and scenario-based training systems

Figure 1, inspired from [8], tries to capture the relationships between competency-based training, evidence-based training as flight data monitoring programs feed in information for training development at all levels, and scenario-based training which constitutes an essential element for providing learner centered experiences. In addition to the closed workflow between A) training goals and objectives; B) competencies, knowledge, and skills; C) tasks; and D) scenarios, Figure 1 distinguishes on the left hand side training development including: the specification of competency frameworks, sociotechnical task analysis, and scenario generation. The right hand side of the figure presents key elements related to the measure and evaluation including: performance measurement, knowledge component assessment, and program evaluation.

The remaining sections of the paper fall essentially within the right hand side of Figure 1 under “Knowledge Component Assessment”. The courseware delivery software gathered the student learning performance data during the learning process, including the sequences of activities selected by the students, timestamps, and question answers.

### 3. DATA DESCRIPTION

The data consists of two sets of web training sessions engaging students on scenarios requiring information gathering, review and assessment of new flight procedures with demands on both knowledge and skill acquisition related to taking off and landing operations. The two data sets correspond to two separate groups of students, and had respectively eight and six students in them. Table 1 presents the frequency distribution of events either as being assessments or information-gathering events for each student in the two groups. The counts in Table 1 refer to the sum of single events. For example, student 1 in Group 1 was assessed 46 times and gathered information 503 times. Essentially, information-gathering events refer to pages containing texts or videos, and assessment events refer to pages where an evaluation of knowledge or skills is performed. Overall the student pilots in the first group had a ratio of about 9% of assessment for information gathering events, while the pilot students in the second group had a ratio of about 13%. The number of assessments includes repeated trials on assessment items. Given that the following sections focus on specific subsets of observations (ex. frequent sequences, or first attempt assessments only), Table 1 provides a high-level view and context for these learning events analysis.

Table 1. Distribution of assessment and information events for each student in the two groups.

Students	Assessment	Information	Total
<b>Group 1</b>			
1	46	503	549
2	45	497	542
3	51	514	565
4	42	495	537
5	52	477	529
6	49	512	561
7	47	547	594
8	57	478	535
<b>Group 1 Total</b>	<b>389</b>	<b>4023</b>	<b>4412</b>
<b>Group 2</b>			
a	42	305	347
b	55	323	378
c	37	259	296
d	34	280	314
e	41	311	352
f	37	284	321
<b>Group 2 Total</b>	<b>246</b>	<b>1762</b>	<b>2008</b>
<b>Grand Total</b>	<b>635</b>	<b>5785</b>	<b>6420</b>

### 4. SEQUENCE MINING

The objective of the application of sequence mining techniques to the learner dataset was to test the hypothesis that students who acted similarly in training would also perform similarly in the assessments. Results indicate that a significant relationship between students’ behavioural patterns during training and performance on test problems exists.

For the analysis in this section, we utilized a data-driven approach to classify student activity and behaviour patterns in the web training courseware, with the purpose of identifying dependencies between the way students interact with the training material, and how the students perform on subsequent assessment-based tests and exercises. At a high level, the working hypothesis for this part of the study is thus that students who behave similarly (i.e. by exhibiting similar patterns of navigation activity when interacting with the courseware) will perform similarly in the assessments.

To test this hypothesis, we classified the students into two groups, using three different criteria: 1) those who scored above the median score on the assessments versus those who scored below the median, 2) those who scored above average on assessments versus those who scored below, and 3) classification according to response similarity. For this final classification scheme, we considered similarities in student success on a question-by-question basis. A distance function was introduced, with the distance between two students defined as the number of assessment questions for which one student gave the correct response and the other gave an incorrect response. K-means clustering was then used to divide the students into two groups in which in-class distances were minimized. Thus two students in the same class were likely to have scored the same (correct or incorrect) more often than two students in different classes. This particular analysis thus more closely strives to validate the working hypothesis that students who behave similarly will perform similarly in the assessments. So, rather than only judging similarity between two students only in terms of total score, we also took a view of how they scored in relation to each other in terms of the number of assessments in which both responded correctly or both responded incorrectly.

For each classification scheme above, the hypothesis is that students classified in the same group (i.e. those whose score similarly in assessments in terms of total score or response similarity) should have exhibited more similarities in how they interacted with the courseware during the learning phase. To test this, we utilized sequential pattern mining (using the SPAM [9] algorithm) to mine sequences of behaviour that were discriminative of each group (i.e. sequences of pages visited that were found to be highly frequent in one group and highly infrequent in the other), and then used leave-one-out cross-validation to test our ability to correctly classify each student based on the existence of these mined behavioural sequences.

Figure 2 shows the accuracy of our classifier for each classification scheme. For example, the leftmost bar indicates that we were able to correctly classify whether a student scored above or below the median score in 93% of the cases (as well as above/below average in 100% of cases and according to response similarity in 86% of cases), solely through analysis of behaviour patterns exhibited by the students when navigating through the courseware. The p-value for each statistic indicates the probability of achieving these results (or better) purely by chance. This indicates that a significant relationship exists between students' behavioural patterns during training and performance on test problems.

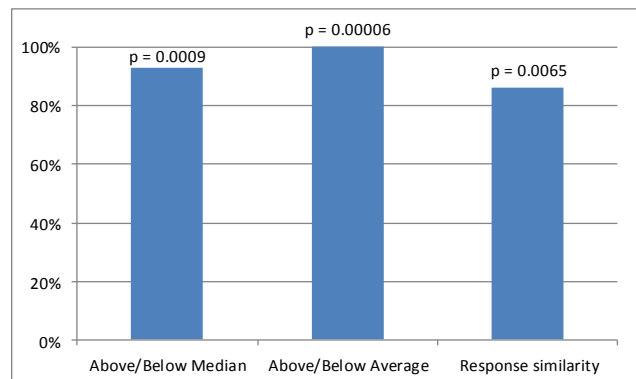


Figure 2. Results of sequence classification on students

To further examine the relationship between behaviour and results, we took a closer examination of the similarities between students when classified as either above or below average score, the scheme that was most successful in the test above. Here we generated the set of frequent behaviour patterns exhibited by each student, and then computed the Jaccard similarity of each pair by quantifying the degree of overlap in the set of frequent patterns for each student, where the Jaccard similarity of two sets A and B is equal to the size of the intersection of A and B, divided by the size of the union. Table 2 summarizes these results by showing, for each student, the average similarity to students who placed above and below the average. On average, students achieving a lower than average score had more similar behaviour to other students who achieved a lower than average score, and vice-versa. In fact, in all cases but one, each student behaved more similarly on average to students in its own group.

Table 2. Average similarity for each student to students with below/above average score

Below Average Students			Above Average Students		
Student	Similarity with below average students	Similarity with above average students	Student	Similarity with below average students	Similarity with above average students
1	0.125	0.080	3	0.059	0.071
2	0.078	0.068	4	0.100	0.075
5	0.047	0.033	a	0.051	0.068
6	0.070	0.061	b	0.063	0.112
7	0.032	0.026	c	0.024	0.042
8	0.127	0.075	d	0.063	0.133
			e	0.040	0.072
			f	0.059	0.142
Average	0.080	0.057		0.057	0.090

While there are wide-ranging behaviours that differentiate the two groups, Figures 3 and 4 point to two interesting behaviour patterns that were particularly prevalent in the initial dataset of 8 students. The first instance, in Figure 3, was highly frequent among the higher-achieving group, and quite infrequent among the lower-achieving group. This behaviour shows a lot of activity reviewing notes before completing a particular section and moving on. This could indicate that this note review had an impact on the success of the students. The second instance, in Figure 4, was highly frequent among the lower-achieving group, and quite infrequent among the higher-achieving group. This behaviour shows a lot of activity around calculations regarding take-off. This could provide

a clue into where the less successful students are going wrong, and thus where improvements to the courseware may be made.

1. Review\_Introduction\_1, Review\_Introduction\_2,
2. Full\_Review\_Notes\_Mission\_Planning\_1,
3. Full\_Review\_Notes\_Landing\_Limits\_and\_Procedures\_2,
4. Full\_Review\_Notes\_Landing\_Crosswinds\_3,
5. Full\_Review\_Notes\_Takeoff\_Procedure\_4,
6. Full\_Review\_Notes\_Takeoff\_Conditions\_5,
7. Full\_Review\_Notes\_Takeoff\_Crosswinds\_6,
8. Full\_Review\_Notes\_Landing\_Calculations\_7,
9. Full\_Review\_Notes\_Takeoff\_Calculations\_8,
10. Full\_Review\_Notes\_ControlUnit\_Invalid\_9,
11. Full\_Review\_Notes\_ControlUnit\_Calculations\_10,
12. Transition\_To\_Test-GUI\_MAP,
13. Lesson\_Conclusion\_Pass

Figure 3. Example behaviour of the higher-performing group

1. Select\_Calculation-Takeoff\_Crosswinds\_1-
2. Select\_Calculation-Takeoff\_Pitch\_1-Takeoff\_Pitch\_2,
3. GUI\_MAP-Calculations\_Introduction\_1-  
Calculations\_Introduction\_2-  
Calculations\_Introduction\_3- Invalid\_11-Invalid\_12,
4. Invalid\_14-How\_To\_Use\_Introduction\_1-  
How\_To\_Use\_Introduction\_2,

Figure 4. Example behaviour of the lower-performing group

This result has a number of implications. First, it demonstrates a tangible correlation between how students choose to navigate the courseware and how well they perform on assessments. Second, it establishes clear evidence that opportunities exist to predict student achievement during the learning phase, when remedial action can be taken to improve comprehension. Finally, the ability to identify the key behaviours that have the highest impact on how a student will perform can facilitate strategic managerial decision making on how to direct the flow of student activity through the courseware.

## 5. ADDITIVE FACTOR MODELS

The Additive Factor Model (AFM) was chosen because it represents a common technique in educational data mining [12]. By using this data analysis technique, we were seeking estimations for parameters for student proficiencies, as well as items difficulty, and competencies easiness. AFM is a model for assessing the quality of an items-to-skills mapping, based on its ability to predict empirical observations of student results [10]. It may be seen as a generalization of Item Response Theory [11], where the response depends not only on item difficulty and student proficiency, but also on underlying knowledge components (KC) and the sequence in which they are met. In AFM, these knowledge components can be associated with competencies, skills, or declarative knowledge that are responsible for a student's performance. The mapping between an item (question, task, problem) and knowledge components is provided in the form of a binary Q-matrix  $\mathbf{Q}=[q_{ik}]$ , where  $q_{ik}=1$  indicates that item  $i$  is associated to knowledge component  $k$  [13]. The probability that a student  $j$  will correctly answer an item  $i$  is modelled using a mixed-effect logistic regression

$$P(Y_{ij} = 1|\alpha, \beta, \gamma) = \frac{1}{1 + \exp(-(\alpha_j + \sum_k \beta_k q_{ik} + \sum_k \gamma_k q_{ik} t_{jk}))} \quad (1)$$

where  $\alpha_j$  is the proficiency of student  $j$  (higher proficiency yields higher success rate),  $\beta_k$  is the easiness and  $\gamma_k$  the learning rate for knowledge component  $k$  (higher easiness yields higher success,

higher learning rate means increased success on subsequent trials)<sup>2</sup>. The observed student sequence is summarized in the opportunity  $t_{jk}$ , i.e. the number of times student  $j$  has met knowledge component  $k$ . As learning progresses, increasing opportunity translates into higher probability of success in items associated with that KC.

Our learner dataset contains 38 items, taken by 14 students (in two sessions of eight and six) between zero and four times each, resulting in 533 transactions.<sup>3</sup> The course designers provided the Q-matrix mapping the 38 items to 14 knowledge components (Figure 5, where the items are the specific questions or problems that the students had to answer or solve, while the knowledge components are the underlying knowledge and skills accounting for the learner's performance on those questions or problems.

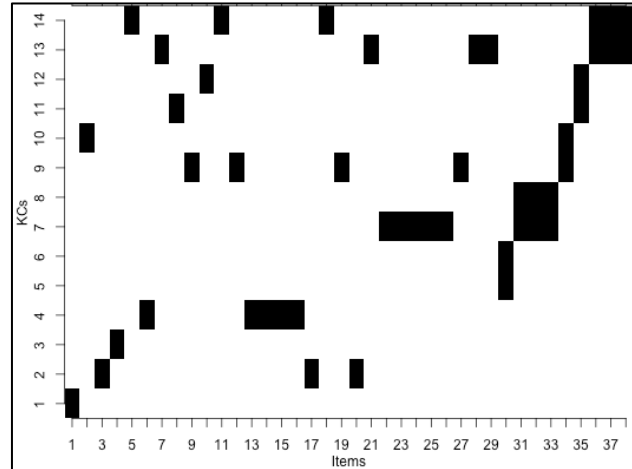


Figure 5: Q-matrix from courseware designer: 38 items x 14 KCs.

Estimation of the AFM model parameters is done by maximizing the likelihood<sup>4</sup> on the transactions, with the constraint that learning rates are kept positive, and a slight regularization on the alpha parameters in order to keep them within the  $[-3; 3]$  range.

### 5.1 Student Proficiency

We analyse the proficiency of the two groups of students using the estimated alpha parameters. Figure 6 shows that the first group of students (1-8) has overall a lower proficiency than the second group (a-f). The two students with lower proficiency in the second group (b and c) have estimated proficiencies on par with the best two students from the first group (3 and 4). Student 5 clearly displays the lowest proficiency by far.

This is partly reflected in the observed success rates, which range from 58.5% for student 5, to 100% for student d. We learned *post analysis* that the second group had received an improved set of instructions. Although there was no difference between the first and second groups in expectations, motivation or engagement with the training material, the improved instructions have a clear

<sup>2</sup> Proficiency and easiness values are relative to the other values in the set, and should not be interpreted as actual success rates.

<sup>3</sup> Each transaction records one student's result on one item.

<sup>4</sup> We use a conjugate gradient algorithm. Any optimization method would work similarly as the log-likelihood is convex.

impact on the estimated proficiency for the second group. This validates the effectiveness of the change.

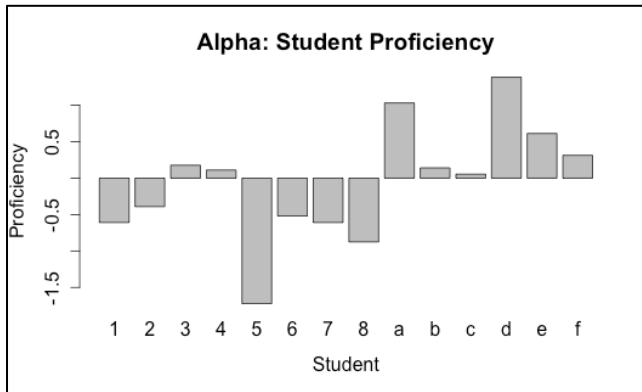


Figure 6: Student proficiency, estimated by AFM.

## 5.2 Competency Analysis

We analyse the competencies through the estimated beta and gamma parameters. Note that the actual parameter values are difficult to interpret separately, as various combinations of beta, gamma and opportunity may yield similar probabilities (Eq. 1). They do make sense in combination of the base “easiness” beta and learning rate gamma, to explain how the probability of success changes as the number of opportunity increases. As a consequence, rather than looking at actual parameter values, we relate them to the corresponding prediction ability. We analyse competencies by looking at the probability to fail on items associated by each knowledge component on the first three opportunities, for a hypothetical student with a proficiency parameter of zero. Figure 7 shows this for 11 knowledge components (The easiest KCs, 1, 4 and 11, get 0% for both predicted and observed error from the first attempts).

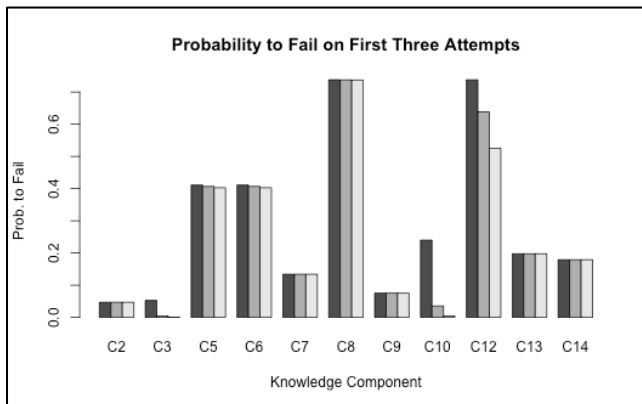


Figure 7: Probability of error for several knowledge components.

Note that due to the constraint that the learning rate is positive the probability to fail is always decreasing (Eq. 1). Learning is clearly apparent for several competencies (C3, C10 and C12), as shown by the clear drop in probability to fail as the KC is addressed. For C5 and C6, learning is much slower, and the error rate stays around 41%. However, this observation should be mitigated by the fact that these knowledge components are only associated with one item and always together (Figure 5). There is therefore very little data to estimate learning on these competencies, as most students took that item only once. When considered in combination in item #30, KCs C5 and C6 yield a predicted error

on this item of 36%. In addition, this points to a possible refinement of the Q-matrix: these two knowledge components could be merged with no loss of modelling capacity.

Probability of failure seems consistently high for C8. However, Figure 5 shows that this knowledge component always appear together with C7 (which also appears alone). Due to the additive nature of the AFM model, the actual probability of success for items featuring C8 actually combine the easiness and learning rates for both C7 and C8, resulting in a probability of failure of 30.3%. Items involving both C7 and C8 are significantly harder than items involving C7 alone, and the AFM model adjusts for this fact by estimating a low easiness (high difficulty) for knowledge component C8.

The analysis of the AFM results therefore provides us with non-trivial insight into 1) the proficiency of the students taking the course, and 2) the difficulty and learning rates of the various competencies addressed in the course. It also suggests possible refinements of the competency framework produced by the course designer. Finally, despite the clear difference between the two groups of students, we have also observed that the estimates for the parameters related to competencies ( $\beta_k$  and  $\gamma_k$ ) are consistent across the two groups.

## 6. CONCLUSION

To address the challenges of pilot training in the early 2000s, civil aviation stakeholders like CASA, ICAO, and concurrently the USAF have been promoting competency-based training as a training model. In addition to focusing on competencies rather than hours, the industry has also brought to bear actual flight monitoring data as a source to determine learning objectives. The essential element evidence-based training introduces to competency based-training is the reference to operational data as a means to identify key competencies, in addition to the analysis performed by subject matter experts. A literature review also revealed that a combination of competency, evidence, and scenario-based training approaches can form the basis for the next generation of pilot training system. The latter approach being consistent with the principles of situated learning theory, which argues that learning best takes place in the context in which it is going to be used. The paper focused essentially on the assessment of knowledge components using sequence mining and logistic regression for the purpose of understanding learning processes and improving learning scenarios. The data used for these analyses was collected in the context of pilot training using a scenario-based approach for reviewing basic landing and taking off flight operations.

The objective of the application of sequence mining techniques to the learner dataset was to test the hypothesis that students who acted similarly in training would also perform similarly in the assessments. Results indicate that a significant relationship between students’ behavioural patterns during training and performance on test problems exists.

The Additive Factor Model, a model for assessing the quality of an items-to-skills mapping based on empirical observations of student results, was used to estimate student proficiency and knowledge components difficulty. Our analysis indicated a clear difference between students from two groups in the data. It also helped us identify competencies that are inherently easy, as well as hard competencies for which learning allows the probability of failure to quickly drop over subsequent attempts. It also suggests changes in the competency framework in which knowledge components could be merged with no loss of modelling capacity.

Together, the application of the descriptive method of sequence mining, and the predictive technique of additive factor models, provide results that may be used to evaluate and improve instructional design.

Some potential future directions for the project include: a) collecting more data, using the same approach for additional data sets, and comparing the result; b) developing alternative methods, and using the methods on same data sets to test and compare results; and c) conducting validation with instructional design experts in the relevant domain.

## 7. ACKNOWLEDGMENTS

The NRC project team would like to thank Dr. Susanne Lajoie (McGill University), who helped the NRC team to obtain its ethics certificate by providing the relevant documentation supporting McGill's ethics request to process CAE pilot learning data. The authors would also like to thank the following reviewers from CAE: Paula Mazzaferro, David Graham, and Graham Estey.

## 8. REFERENCES

- [1] International Civil Aviation Organization, *Manual of Evidence-based Training*, First edit. Montreal, Canada: International Civil Aviation Organization, 2013.
- [2] Civil Aviation Safety Authority, "Competency Based Training and Assessment in the Aviation Environment," 2009.
- [3] C. M. Colegrove and G. M. Alliger, "Mission Essential Competencies: Defining Combat Mission Readiness in a Novel Way," in *RTO SAS Symposium on "Air Mission Training Through Distributed Simulation (MTDS) Achieving and Maintaining Readiness,"* 2002, vol. 323, p. 22.
- [4] International Civil Aviation Organization, *Quality Assurance Manual for Flight Procedure Design. Flight Validation Pilot Training and Evaluation (Development of a Flight Validation Pilot Training Programme)*, First edit., vol. 6. Montreal, Canada: International Civil Aviation Organization, 2012.
- [5] R. Chapman and C. Colegrove, "Transforming operational training in the Combat Air Forces.," *Mil. Psychol.*, vol. 25, no. 3, pp. 177–190, 2013.
- [6] Air and Space Academy, "Dealing with Unforeseen Situations in Flight," Bruguieres, France, 2013.
- [7] M. Thurber, "The future of pilot training," *Aviation International News Online*, 2014. [Online]. Available: [http://www.flightresearch.com/pdfs/AIN\\_Pg\\_20-28.pdf](http://www.flightresearch.com/pdfs/AIN_Pg_20-28.pdf). [Accessed: 01-Feb-2015].
- [8] J. MacMillan, E. B. Entin, R. Morley, and W. Bennett, "Measuring team performance in complex and dynamic military environments: The SPOTLITE method.," *Mil. Psychol.*, vol. 25, no. 3, pp. 266–279, 2013.
- [9] J. Ayres, J. Flannick, J. Gehrke, and T. Yiu, "Sequential pattern mining using a bitmap representation," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2002, pp. 429–435.
- [10] H. Cen, "Generalized Learning Factors Analysis: Improving cognitive Models with Machine Learning," Carnegie Mellon University, 2009.
- [11] R. D. Bock, "A Brief History of Item Theory Response.," *Educ. Meas. Issues Pract.*, vol. 16, no. 4, pp. 21–33, 1997.
- [12] A. Pena-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Syst. Appl.*, vol. 41, no. 4 PART 1, pp. 1432–1462, 2014.
- [13] K. K. Tatsuoka, "Rule space: an approach for dealing with misconceptions based on item response theory," *J. Educ. Meas.*, vol. 20, no. 4, pp. 345–354, 1983.