Defence Research and Development Canada
Recherche et développement pour la défense Canada

DRDC | RDDC
technologyscienceetechnologie

Learning Retention Effects and Student Engagement for a Basic Aviation Medicine Course

A Pilot Study of Baseline Performance

Blake C. W. Martin
DRDC – Toronto Research Centre

The body of this CAN UNCLASSIFIED document does not contain the required security banners according to DND security standards. However, it must be treated as CAN UNCLASSIFIED and protected appropriately based on the terms and conditions specified on the covering page.

Defence Research and Development Canada
Scientific Report
DRDC-RDDC-2020-R041
April 2020

Canada

IMPORTANT INFORMATIVE STATEMENTS

This document was reviewed for Controlled Goods by Defence Research and Development Canada (DRDC) using the Schedule to the *Defence Production Act*.

Disclaimer: This publication was prepared by Defence Research and Development Canada an agency of the Department of National Defence. The information contained in this publication has been derived and determined through best practice and adherence to the highest standards of responsible conduct of scientific research. This information is intended for the use of the Department of National Defence, the Canadian Armed Forces ("Canada") and Public Safety partners and, as permitted, may be shared with academia, industry, Canada's allies, and the public ("Third Parties"). Any use by, or any reliance on or decisions made based on this publication by Third Parties, are done at their own risk and responsibility. Canada does not assume any liability for any damages or losses which may arise from any use of, or reliance on, the publication.

The data collected as part of this study was approved either by Defence Research and Development Canada's Human Research Ethics Board or by the Director General Military Personnel Research & Analysis' Social Science Research Review Board.

Endorsement statement: This publication has been peer-reviewed and published by the Editorial Office of Defence Research and Development Canada, an agency of the Department of National Defence of Canada. Inquiries can be sent to: Publications.DRDC-RDDC@drdc-rddc.gc.ca.

© Her Majesty the Queen in Right of Canada (Department of National Defence), 2020

© Sa Majesté la Reine en droit du Canada (Ministère de la Défense nationale), 2020

Abstract

Classroom technologies such as SmartBoards, collaboration software and hardware and classroom response systems are seen as enablers in the effort to modernize individual training and education (IT&E). It is impossible to verify and validate the efficacy of such tools without first establishing baseline measures of student performance in courses, and engagement with the learning experience. Most tests in the Canadian Forces are pass/fail criterion-referenced, and do not provide sufficiently granular data to understand the influence of specific interventions. Using norm-referenced measurement, we tested students on how much they knew about aviation medicine before and after the Basic Aviation Medicine (BAvMed) course offered through the School of Operational Medicine (SOM) within the Canadian Forces Environmental Medicine Establishment (CFEME). Students performed a test that paralleled the content and structure of the qualification test at one week before, immediately after, six weeks after, and six months after the course. Test scores improved by 13 more correctly answered questions from before to after the course, and we observed very little decay after six months. We also compared test scores from the parallel and qualification tests and found there was no difference between the two, suggesting that the parallel test is a reasonable measure of student achievement for this course. Last, a questionnaire on student engagement revealed no correlation between test scores and any measure of engagement. These results provide performance benchmarks and methodological validation for future investigations of educational technology in the context of the BAvMed course.

Significance to Defence and Security

In contrast to pass/fail rates, the amount students learn during training and the amount they retain afterward provide meaningful metrics for the success of a given course, and permits the alteration, variation and assessment of training materials, methods and technology. Such baseline information will allow the Canadian Armed Forces (CAF) to appropriately apply resources within training programs and ultimately improve operational effectiveness.

Résumé

Les technologies en classe, comme le tableau numérique interactif ou smartboard, le matériel et les logiciels de groupe, ainsi que les systèmes d'intervention pédagogique sont perçus comme des outils dans les efforts de modernisation de l'instruction individuelle et l'éducation (II et E). Il est impossible de vérifier et de valider l'efficacité de ces outils sans avoir d'abord établi une base de référence pour évaluer le rendement des étudiants en classe et leur engagement à l'égard de l'expérience d'apprentissage. La plupart des tests dans les Forces armées canadiennes fonctionnent selon le critère réussite ou échec et ne fournissent donc pas de données suffisamment détaillées pour nous permettre de comprendre l'influence d'une intervention en particulier. Au moyen de mesures normatives, nous avons évalué les connaissances des étudiants en matière de médecine aéronautique avant et après avoir suivi le cours élémentaire de médecine aéronautique offert par le Centre de médecine environnementale des Forces canadiennes (CMEFC). Les étudiants se sont soumis à un test dont le contenu et la structure sont analogues à ceux du test de qualification une semaine, immédiatement après, six semaines après et six mois après le cours. Les étudiants ont répondu correctement à 13 questions de plus après le cours, comparativement au premier test, et nous avons constaté très peu d'érosion des connaissances après six mois. Nous avons également comparé les résultats du test analogue au test de qualification et nous n'avons trouvé aucune différence entre les deux, ce qui suggère que le test analogue est une mesure raisonnable du rendement des étudiants de ce cours. Finalement, un questionnaire sur l'engagement des étudiants n'a révélé aucune corrélation entre les résultats aux tests et leur degré d'engagement. Ces résultats fournissent des seuils de référence pour le rendement et une validation méthodologique pour de futures enquêtes sur les technologies éducatives dans le cadre du cours élémentaire de médecine aéronautique.

Importance pour la défense et la sécurité

En comparaison du critère réussite ou échec, la mesure de la quantité de matière que les étudiants apprennent pendant le cours et retiennent par la suite est utile pour déterminer la réussite d'un cours donné et permet de modifier, de varier et d'évaluer la documentation, les méthodes et les technologies de la formation. Ces renseignements de base permettront aux Forces armées canadiennes (FAC) d'utiliser de manière appropriée les ressources dans les programmes de formation et, au bout du compte, d'améliorer l'efficacité opérationnelle.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Thanks to the School of Operational Medicine for allowing access to their students and courses and to the Information Technology group at Defence Research and Development Canada (DRDC) – Toronto Research Centre for their generous assistance in data collection, and Dr. Stuart Grant for his guidance on this project. A special thank you to Lieutenant Colonel Anna Grodecki and Major Rachel Morrell for their help and expertise during development of experimental materials, and for their insightful comments on the manuscript.

1 Introduction

1.1 Training Modernization

The Canadian Armed Forces (CAF) is exploring ways to modernize and streamline its delivery of Individual Training and Education (IT&E) [1–4]. As such, efforts at modernisation are underway, including use of instructional technologies and novel methods. Until we understand the success of current courses and methods, it will be difficult to develop a critical understanding of how interventions impact learning and retention. The School of Operational Medicine (SOM) offers advanced training in aerospace medicine and hyperbaric diving medicine and is part of the Canadian Forces Environmental Medical Establishment which shares facilities with DRDC – Toronto Research Centre. SOM courses are typically offered in a lecture and case study format, and may be amenable to delivery improvements through application of novel instructional technologies.

1.2 Measures of Learning Success

One possible assessment of learning success is a summative achievement test [5]. It is possible to further categorize such tests by whether the result is compared against a minimum performance standard, in which case it is called a 'criterion-referenced' test, or against the performance of a given population, in which case it is called a 'norm-referenced' test. In criterion-referenced testing, students are said to pass if they meet or surpass a cut-off for a specific behaviour or knowledge of interest [6]. In other words a student passes if they possess the behaviour in sufficient measure, and rather than report a grade, the degree of success is reported simply as a pass or fail. By contrast, norm-referenced tests use the knowledge of a given population as the basis of comparison for success with respect to a particular body of knowledge; in other words, students pass if they meet a cut-off (typically 50% correct for most academic courses), with the key concept being that students are compared to each other within the population, with some performing better or worse than their peers [7].

The advantage of a criterion-referenced test for the CAF is that it ensures a known level of competence for a given domain [8]. In the instance where such tests are recorded as pass or fail, the disadvantage is that it does not facilitate comparison among members of the population. Also in some courses CAF members are permitted more than one attempt at a summative test, with feedback between attempts as a means of further developing the student. As a result pass/fail rates do not necessarily reflect the efficacy of the instructional design for a given course. Moreover, the standard of success is frequently a series of complex behaviours or an integrated body of knowledge, so the criterion-referenced test does not provide adequate granularity to compare instances of learning among individuals or the impact of a given learning intervention. Therefore it is important for research purposes to use separate norm-referenced tests, while the criterion-referenced test for the course remains the benchmark for success within each participant's career progression. Notwithstanding the above rationale for criterion-referenced testing, pass-fail data do not provide sufficient granularity to determine which content is required, or the degree to which an intervention, technology or teaching approach can be correlated with student success on a course. Further, norm-referenced testing allow for examination of the distribution of responses among in the representative sample, which can provide additional information about the competence of the group in question, and provide additional information relevant to course formation and design [7, 8].

In the case of the SOM, students must meet a minimum criterion-referenced standard (70%) for the summative theoretical test; however, the nature of the test for the BAvMed course provides a measure that can also be referenced against the norm. In the present experiment this gave us the opportunity to compare a norm-reference test that paralleled the structure of the qualification test.

In some instances, the degree to which students are involved in the learning process (student engagement) is taken as a measure of success [9]. Studies concerned with technology use often explore student engagement; however this measure does not always correlate with test scores [10].

1.3 Testing Effect

A test with content that 'parallels' content in the official tests can prevent familiarization and memorization of specific questions and content in the criterial test, preserving value of the official test. It is not possible to eliminate the testing effect however, and some study participants can be expected to perform better on the official test than those who did not participate. The one caveat with respect to testing and retesting participants is the 'testing effect': that is students who are repeatedly tested on the same material show gains in learning and retention, even when they are not provided with feedback on the test [11] and even when parallel tests are used to explore the same material [11, 12].

1.4 Pilot Study

We performed a pilot study on a group of students from a Basic Aviation Medicine (BAvMed) course in the Canadian Forces Environmental Medicine program. The learning takes place over 19 days at the SOM in Toronto. The course currently uses a traditional, lecture style format presented on PowerPoint, and covers a wide variety of topics including ophthalmology, addiction, internal medicine, physiology, neurology and many others. Course content is developed under the provisions of the Canadian Forces Individual Training and Education System (CFITES) policy. A qualification standard (QS) for the population in question gives rise to a training plan (TP), which in turn suggests means and tools for assessment to determine achievement of a given performance objective (PO) [13]. Based on the CFITES framework, SOM courses are delivered by fully qualified instructors who design educational experiences to address learner needs based on the current literature and input from medical specialists. In some instances specialists will appear as guest teachers either in person or by telepresence.

This study investigated learning, retention of learning, and student engagement in a specific CAF course in the School of Operational Medicine. In addition to the normal assessment tests for the course, we used an online comprehensive parallel test to measure pre-existing knowledge about the course subject matter, and provide a reference criterion for future studies. We also used a student engagement questionnaire to determine if academic success was correlated with the degree to which students felt involved with their learning.

2 Method

2.1 Participants

Study participants volunteered from a course at the SOM. There were 25 students in the course, with seven females. All participants were fully qualified clinical practitioners (general duty medical officers or physician's assistants) with training applicable to their trade (basic medical officer training). A total of 24 students volunteered to participate in the study. All 24 performed the Pre-Test, 16 performed the Post-Test, 10 performed the six week retention test and five performed the six month retention test. The demographic composition of subjects was not available.

A control group of seven clinical practitioners (general duty medical officers who are eligible to take the BAvMed course) were solicited from a closed Facebook user group. They performed the Pre-Test and Post-Test 19 days later without attending classes or engaging with the learning material. They were instructed to not study nor prepare for the tests in any way.

All participants gave informed consent, and testing was conducted in accordance with provisions of DRDC's Human Ethics Review Board, Protocol 2014-043.

2.2 Procedure

Prior to the course, the author and two directors of the BAvMed course developed a series of multiple-choice questions that probe knowledge, understanding and application of information that is germane to the course. These questions were not part of the question bank for exams within the course, but were developed with a view to eventual inclusion. In other words, these questions were designed to fairly test information students are expected to know. For the parallel test there were 36 questions total with 45 correct answers out of 165 possible answers (Annex A). The qualification test had 37 questions total, with 47 correct answers out of 160 possible answers.

Students (intervention group) were provided with information about the study, and gave informed consent after volunteering to participate. Participants engaged in all aspects of the course including tests and homework. Participants performed the parallel test (Pre-test) prior to commencement of the course, and again immediately following the completion of the course (Post1), after the qualification test. The time between the Pre-test and Post1 was 19 days. An Early Retention test was performed six weeks (Post2) following the qualification test, and a Late Retention test performed after six months (Post3).

Since the qualification course for the test was performed on paper, and we intended to administer the 6-week and 6-month retention courses on computers, we tested to see if performance differed between paper and online versions of the test. Half of the volunteers were randomly selected to perform the test with paper and pencil, and the other half responded on an online, computerized version of the test that was offered securely using the using the FluidSurveys online survey software (http://fluidsurveys.com).

Following the Post1 test, the intervention students also completed the Student Course Engagement Questionnaire (SCEQ) [14] (Annex B). We slightly modified some wording to correspond more closely with a military learning environment. The questionnaire is a 23 item, 5 interval Likert scale type

inventory, consisting of Skills Engagement, Emotional Engagement, Participation Engagement, and Performance Engagement subscales as well as a Total Engagement measure.

Participants in the control group performed the Pre-Test online, and then performed the same test (Post1) 19 days later with no instruction between the two tests.

2.3 Data Analysis

An 'item analysis' was used as the principal means of measuring differences in performance. In such testing, questions are used as the unit of analysis, and success on each question is compared among test instances. Typically, such item analyzes are used to validate questions within a test, and the test as a whole, particularly with multiple choice questionnaires [15]. We chose this analysis as a way to mitigate undesired statistical effects of unequal group size between the intervention and control groups, diminishing size of the retest group, and possible re-testing bias. Changes in the performance of individual questions could then be reliably analyzed. Additional t-tests and correlational analyzes were performed to examine relationships between the groups of students, and relationship of student performance to student engagement.

Some questions had more than one possible correct answer. In this case, all correct answers had to be selected for the question to be counted as correct. Separate χ^2 tests were performed on each question for three analyzes: 1) Learning, before and after the 19-day BAvMed course (Pre, Post1); 2) Control, before and after a 19-day period with no course (Control1, Control2); and 3) Retention, immediately, six weeks and six months after (Post1, Post2, Post3). When expected counts were less than five, we applied a Fischer's Exact test.

For the computer input method, it was not possible to omit answers, however for the paper input method, on question 14, ten participants failed to enter the correct number of responses, and on question 24, three participants did not enter enough answers (five selections were required for each, but four or fewer were entered). Omitted answers were marked as incorrect. Regardless, there was no difference between paper and computer input methods for the observed number of correct answers for either question; χ^2 (1, N=40)=.902, p=.342 and χ^2 (1, N=40)=.000, p=1, respectively. We compared input methods using an unpaired t-test for all answers to all questions, finding no difference between paper and computer input methods. (t=.79, df=38, p=.43), and we pooled these data for all subsequent analyzes.

To provide a measure of the validity of the experimental test, the total of all correct answers was tabulated for each Post1 (experimental) test and compared against the total of all correct answers for each qualification test in an unpaired t-test.

Equally, to explore the homogeneity of the intervention (those who took the course) sample in comparison to the control group, we performed an unpaired t-test on the total of all completely correct answers for both groups.

We performed bivariate correlational analyzes between individual test scores for each participant, and that participant's SCEQ score, exploring results for each item, subscale and total score on the inventory.

3 Results

3.1 Learning

We compared performance on each question as answered before (Pre) and immediately after (Post1) the course. In comparing the Pre and Post1 score observations, χ^2 tests determined 13 of 36 questions were more often answered correctly following the course. In five of these instances, expected observations were lower than five, and Fischer's Exact test was used. The results are detailed in Table 1.

Table 1: *Improvement in question performance from Pre to Post1. Where expected observation of five was not met, Fischer's Exact test was used. The accepted value in each instance is marked with an asterisk.*

Quest.	Incor. Pre	Cor. Pre	Incor. Post1	Cor. Post1	χ^2	df	sig	Fischer's Exact
q1	6	4	3	13	4.626	1	.031	.046*
q2	6	4	2	14	6.518	1	.011	.025*
q6	18	6	4	12	9.697	1	.002*	.003
q8	7	17	0	16	5.657	1	.017	.029*
q9	15	9	3	13	7.424	1	.006*	.01
q15	11	13	2	4	4.862	1	.027*	.04
q16	23	1	9	7	9.401	1	.002*	.004
q18	8	16	0	16	6.667	1	.010	.013*
q23	18	6	0	16	21.818	1	.000*	.000
q32	17	7	1	15	16.178	1	.000*	.000
q33	17	7	2	14	13.099	1	.000*	.000
q35	9	15	0	16	7.742	1	.005	.006*
q36	15	9	0	16	16.00	1	.000*	.000

3.2 Retention

Using separate χ^2 tests in our comparison of each question immediately after, six weeks after and six months after, we observed improvements after six weeks on questions 14 and 27, and decrements on questions 16 and 34, listed below in Table 2.

Table 2: *Changes in performance of individual questions for post-tests. The χ^2 tests compare changes in performance between Post1 and Post2, and between Post2 and Post3.*

Quest.	Incor. Post1	Cor. Post1	Incor. Post2	Cor. Post2	Incor. Post3	Cor. Post3	χ^2	df	sig
q14	7	9	0	10	0	6	8.96	2	.011
q16	9	7	9	1	6	0	6.200	2	.045
q27	5	11	0	10	0	6	5.926	2	.05
q34	6	10	8	2	5	1	6.365	2	.041

These results are summarized in Table 3, showing the mean number of correct answers for all participants at each time point, and in Table 4 showing the change in the number of questions answered correctly in comparison with the previous test.

Table 3: *Mean score as a function of test (score out of 36).*

	Pre	**Post1**	**Post2 (6 wks.)**	**Post3 (6 mos.)**
Correct	15.58 (3.4)	24.75 (3.0)	23.6 (2.7)	23.5 (1.0)
Incorrect	20.41	11.25	12.4	12.5

Table 4: *Change between each test in number of questions answered correctly. The Pre-test was used as baseline for Post1.*

	Learning			
	Pre	**Post1**	**Post2**	**Post3**
Correct	*baseline*	13	2	no change
Incorrect	*baseline*	no change	2	no change
		Retention		

Descriptive statistics for the total score of all correct answers show a mean correct percentage of 61% (SD 9.3) with a minimum of 37.8 and maximum of 82.0 for Pre; 79% (SD 6.6), with a minimum of 68.9 and 93.3 for Post1; 77% (SD 6.0) with a minimum of 66.7 and 88.9 for Post2; and 77% (SD 1.2) with a minimum of 75.6 and 93.3 for Post3. There does not appear to be either floor or ceiling effects for the testing instrument (Figure 1).

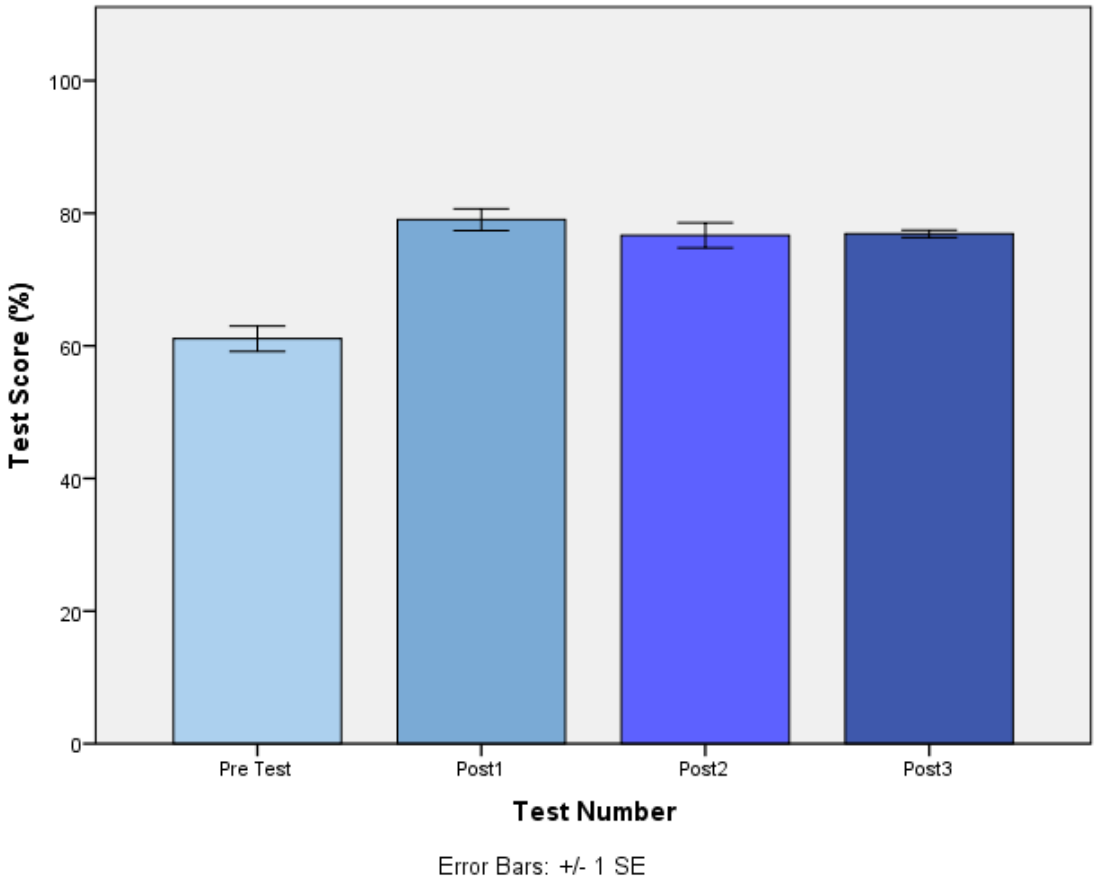


Figure 1: *Mean scores on the parallel test measuring performance before the course, immediately after the course, and six weeks following the end of the course. Student scores improved after taking the course.*

3.3 Controls

3.3.1 Control1 vs. Control2

The χ^2 tests did not reveal any significant improvement for any question for the control group between the first test and the second, 19 days later.

3.3.2 Intervention Pre vs. Control1

In exploring the difference between the control and intervention group, we wished to know whether the two groups were roughly comparable in their initial knowledge and understanding of the material. The intervention group had a mean score out of 36 questions of 14.64 (SD 3.12) completely correct answers in comparison to 16.9 (SD 3.51) for the control group. These scores were not significantly different (t=-1.66, df=22, p=.112).

3.3.3 Experimental Test vs. Qualification Test

An unpaired t-test between the parallel test (mean score 79.02%, SD 6.59), and the qualification test (mean score 77.00%, SD 6.87) showed no significant difference between scores.

3.4 SCEQ

We performed bivariate correlational analyzes between subject Post1, Post2 and Post3 scores and each item, each subscale, and the total scale for the SCEQ. Only cases with valid pairs for a given test were used. We applied a 1000 sample bootstrap procedure at a 95% confidence interval, and found no significant correlations between test scores and student engagement as measured by SCEQ.

4 Discussion

The difference between the number of correct questions observed in the Pre and Post1 tests for the intervention group indicates the BAvMed course successfully teaches students. A lack of improvement in the control group from Control1 to Control2 suggests that the improvement in the intervention group is not simply a result of taking the test more than once. In this instance, it appears that the testing effect did not provide any benefit, most likely because the subjects had no exposure to the material at all, so there were no opportunities for effortful retrieval of the pertinent information after the first test [11]. The performance improved by 13 more questions being correctly answered after the course. These scores may provide a meaningful target for course improvements. Novel combinations of technology and educational approaches should demonstrate an improvement over these values if they are to be deemed worthwhile.

The Pre results of both the intervention and control groups shows that students who meet the qualification to participate in the BAvMed course were able to perform well above chance even before being exposed to the course material (61%). It is important to remember that although a "pass" is 50% on most tests, the midway mark does not represent chance performance when there are more than two choices in a series of multiple choice questions, and further, in tests such as the medicine course, the qualifying grade is 70%. In both tests the students may have been able deduce the answers to some questions because of their knowledge of basic medical principles, or perhaps they already knew the answers to some of the questions. A more detailed analysis of which questions were answered correctly on the Pre Test might provide insight into this line of inquiry and allow trainers to govern the amount of material required for the course. One could discover if there was information that was already possessed by all or most participants, and thus determine if it is worth including.

Thirteen questions showed improved performance after students had taken the course (Table 1). Certain questions were well-performed across each instance of the test. Questions that were well-performed during the Pre-test and subsequent instances could indicate knowledge areas that are unnecessary to cover on the course, or that can be deduced from a student's existing knowledge base. In such a case, it may be beneficial to refocus the course onto material that is more difficult to learn and retain. In the instance of questions where there was improvement between testing instances, the degree of improvement and retention might illustrate subject areas whose content and presentation requires improvement or could indicate a question that was poorly worded or did not adequately assess the learning. For example, question 16, about the vestibular system, was very poorly performed on the pretest (23 incorrect, 1 correct). The performance improved to 9 incorrect and 7 correct for Post1, but then fell to 9 incorrect and 1 correct at six weeks (Post2), and 5 incorrect and 0 correct at six months (Post3). The question is at a knowledge level in Bloom's Taxonomy [16], and is a matter of simple recall. The brief improvement suggests that the information was temporarily available to some students, but was difficult to retain. If the question had been too obscurely worded, there would likely have been no improvement at all from the Pre-test to Post1. Course designers can use this information to examine content and delivery of material related to the vestibular system, or to investigate the relevance and utility of that part of the course. Equally, the structure of the question could be revised in instances where it is insufficiently clear.

These results so far also indicate that students are retaining the information from the course, even though many would not yet have had the opportunity in a clinical setting to employ some or most of the information supplied in the course. Retention scores may indicate that the course is effective in promoting retention, that the students for this particular subject matter and course are likely to retain the information,

that the testing effect promoted recall of the materials, or some combination of these variables. It is likely for the Late Retention (six months) that the reduced variability (SD = 1.2) is a consequence of the lower *n*. Further, the students who actually completed the six month test (Post3) may have been more diligent, and the consequent level of retention on the late test might be attributable to those particular students rather than the course, however correlation of Post3 test results with SCEQ scores does not indicate this is the case. Additional experimentation is necessary to explore these possibilities.

In general, the lack of correlation between test scores and engagement may indicate several things. It is possible that the students in this instance are at a ceiling for measures of engagement. The SCEQ was developed and validated on students in entry level university mathematics classes [14], and it is possible that the attitudes of the BAvMed course students do not fall within the expected behaviour for that instrument either because of their specialized character traits as military personnel, or their specialized character traits as medical personnel. In a recent study with combat arms soldiers, the author used the Motivated Strategies for Learning Questionnaire (MSLQ), and found many correlations among test scores and survey items [17], so the MSLQ may be a better instrument for military populations. Equally, other research has demonstrated success does not always correlate with measures of engagement. This is particularly true for higher achieving students [10]. Testing engagement with an instrument validated on medical students or professionals may clarify some of these questions.

5 Conclusion

The number of questions correctly answered improved by 13 out of 36 for students participating in the BAvMed course, while no improvement in question performance was seen for similarly-qualified controls who did not take the course. It is possible that some of this gain is caused by the testing effect, however without the course, retesting does not provide any benefit. Overall, questions were equally well-performed even after six months. The course appears to provide stable learning with a meaningful effect size, without data crowding at the floor or ceiling, so that either improvements or decrements to course delivery would be evident using a similar experimental model.

Using norm-referenced testing identified particular questions and possibly topic areas that are both well-performed, and poorly performed. Using this information, course designers could make refinements to the questions and their related subject matter presentation within the course to optimize on the use of course time. Additionally, norm-referenced testing provided more finely grained information that was more easily amenable to additional statistical analysis techniques.

A different measure of student engagement may show correlations with test scores. However, a continued lack of correlation may indicate that the students in these courses may be resistant to the Hawthorne effect of enhanced learning simply by virtue of shiny technology [18]. In this way they may provide useful subjects for ongoing explorations of the development of instructional technology and practice. The current course format may meet IT&E goals for training effectiveness, however it is only through careful and ongoing consideration of the delivery of instructional materials that significant and valuable changes can be implemented. We hope to advance the work through a different experimental design and examine the progress of individual students in addition to the testing instruments.

References

[1] Canadian Armed Forces. *NEW* System of Training and Operational Readiness Modernization, http://www.forces.gc.ca/en/business-defence-acquisition-guide-2015/naval-systems-950.page. (2018). Date accessed: October 22, 2019.

[2] Canadian Armed Forces. Evaluation of the Canadian Defence Academy. Cited 2020; Available from: https://www.canada.ca/en/department-national-defence/corporate/reports-publications/audit-evaluation/evaluation-canadian-defence-academy.html. (2015). Date accessed: October 22, 2019.

[3] Canadian Armed Forces. Department of National Defence - Departmental Performance Report 2013-14, https://www.canada.ca/en/department-national-defence/corporate/reports-publications/departmental-performance/2013-14.html. (2014). Date accessed: October 22, 2019.

[4] Miller, D. Individual Training & Education Modernization In The Canadian Armed Forces. *Journal of Advanced Distributed Learning* Technology 1, 47–54, (2013).

[5] National Defence. Manual of Individual Training and Education, Volume 5: Development of instructional programmes. Ottawa, Canada. (1998).

[6] Reiser, R. A. A history of instructional design and technology: Part II: A history of instructional design. *Educational technology research and development* 49, 57–67, (2001).

[7] Klein, T. W. Characteristics Which Differentiate Criterion-Referenced from Norm-Referenced Tests. Nevada State Dept. of Education: Carson City, NV. (1990).

[8] Ornstein, A. C. Norm-referenced and criterion-referenced tests: An overview. *NASSP Bulletin* 77, 28–39, (1993).

[9] Axelson, R. D. & Flick, A. Defining student engagement. *Change: The Magazine of Higher Learning* 43, 38–43, (2010).

[10] Carini, R. M., Kuh, G. D. & Klein, S. P. Student engagement and student learning: Testing the linkages*. *Research in higher education* 47, 1–32, (2006).

[11] Roediger, H. L. & Karpicke, J. D. The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science* 1, 181–210, (2006).

[12] Pagliarulo, C. L. Testing effect and complex comprehension in a large introductory undergraduate biology course. The University of Arizona: Tucson, AZ. (2011).

[13] National Defence. Manual of Individual Training and Education, Volume 4: Design of instructional programmes. Ottawa, Canada. (1998).

[14] Handelsman, M. M., Briggs, W. L., Sullivan, N. & Towler, A. A measure of college student course engagement. *The Journal of Educational Research* 98, 184–192, (2005).

[15] Mehta, G. & Mokhasi, V. Item analysis of multiple choice questions-an assessment of the assessment tool. *Int J Health Sci Res* 4, 197–202, (2014).

[16] Bloom, B. S. Taxonomy of educational objectives. Vol. 1: Cognitive domain. New York: McKay, (1956).

[17] Martin, B. C. W. Obstacle course activity in soldiers impedes recall but not learning of a read text. *Military Psychology*, in press.

[18] Dutta, A., Lee, H. L. & Whang, S. RFID and operations management: technology, value, and incentives. *Production and Operations Management* 16, 646–655, (2007).

Annex A Experimental Test

Aviation Medicine Basic Learning Study

Personal Code: Two letters, two digits: ___ ___ ___ ___

1. At what altitude is breathing supplemental oxygen required to prevent observable adverse physiological effects?:

○ a) 10,000 ft

○ b) 18,000 ft

○ c) 35,000 ft

○ d) 63,000 ft

2. A party balloon placed outside in the hot sun will eventually pop. This is an example of which gas law in action?

○ a) Charles' Law

○ b) Boyles' Law

○ c) Henry's Law

○ d) Dalton's Law

3. What is the tissue level effect of a shift in the oxy-hemoglobin dissociation curve to the right?

○ a) Decreased oxygen delivery to the tissue

○ b) Decreased pH

○ c) Increased oxygen delivery to the tissue

○ d) Decreased gas exchange in the lungs

4. What does a patient with emphysema have in common with a passenger in an unpressurized aircraft at 18,000 ft?

○ a) They are both experiencing hypoxic hypoxia

○ b) They are both experiencing partial atelectasis

○ c) They are both experiencing histotoxic hypoxia

○ d) They are both experiencing hypemic hypoxia

5. The symptoms of middle ear barotrauma in aviation generally occur on _____; these symptoms are a function of _____ Law.

○ a) Ascent; Boyle's

○ b) Descent; Charles'

○ c) Descent; Boyle's

○ d) Ascent; Charles'

6. Which of the following acceleration profiles might result in the push-pull effect?

○ a) +1 Gz followed by + 3 Gz

○ b) – 1 Gz followed by +3Gz

○ c) +3 Gz followed by +1 Gz

○ d) +3 Gz followed by -1 Gz

7. A fighter pilot complains of substernal discomfort with deep inspiration and a dry cough. CXR reveals basal lung collapse with loss of the costophrenic and cardiophrenic angles. You suspect acceleration atelectasis. What features of his recent flight history are a risk for acceleration atelectasis?

○ a) Breathing 100% oxygen

○ b) Flying with an anti-G suit

○ c) Exposure to sustained +Gz acceleration

○ d) All of the above

8. In what way does immersion affect the insulation effectiveness of garments?

○ a) Decreases; water has higher thermal conductivity and heat capacity than air

○ b) Decreases; swimming effort causes sweating, cooling the person from within the garment

○ c) Increases; convective heat loss is lower in water because there's no breeze

○ d) Increases; compression of garments under immersion increases their density, and therefore insulating capacity

9. Why might a pilot in a highly vibration damped seat experience visual difficulty in turbulent flight?

○ a) The instruments should be vibration damped as well

○ b) The pilot is exposed to different vibrations than the aircraft

○ c) The pilot and aircraft are exposed to the same vibration but their responses are different in magnitude and phase

○ d) The pilot would not experience visual difficulty because the vibrations are being damped

10. A positive beryllium lymphocyte proliferation test is evidence that the individual tested has developed:

○ a) Acute beryllium disease

○ b) Chronic beryllium disease

○ c) Sensitization to beryllium

○ d) Beryllium induced metal fume fever

11. A planned period of anchor sleep may aid in fatigue management during periods of demanding work schedules with minimal rest. Anchor sleep:

○ a) Must be a regular sleep period of at least 4 hours duration at the same time each day

○ b) Must be scheduled nap periods of at least 90 minutes throughout the day

○ c) Must be schedule nap periods of at least 60 minutes throughout the day

○ d) Must be obtained by prolonged rest periods before entering an up-tempo schedule

12. A pilot with no previous history of motion sickness presents himself with vertigo and nausea with activities of daily living. The symptoms began after a scuba diving vacation, with minor pain deep in the left ear. The vertigo is most likely caused by:

○ a) Poor adaptation of the visual system following exposure to the underwater environment

○ b) Poor reintegration of visual and vestibular systems after exposure to a pressurized environment

○ c) Possible barotrauma to the vestibular system during the dive

○ d) Increased anxiety about flying following a period of rest

13. Scopolamine is the most effective pharmacological intervention for motion sickness but is not recommended for air crews because:

○ a) Spatial disorientation, sedation, reduced eye-hand coordination and poor reaction time

○ b) Lowered alertness, blurred vision, impaired attention, reduced short-term memory

○ c) Relaxation of voluntary muscle and frequent dosing requirements

○ d) Increased risk of ventricular arrhythmias

14. In general navigators suffer from motion sickness more often than pilots because (select 2):

☐ a) Pilots are selected based on their lower susceptibility to motion sickness

☐ b) Navigators do not have direct visual information about the movement of the aircraft

☐ c) Navigators are oriented sideways to the direction of travel

☐ d) Navigators work in a darker environment causing increased conflict between the visual and vestibular systems

☐ e) Navigators are located closer to the center of the rotation of the aircraft and therefore experience more subtle shifts in perception

15. A search and rescue team member using night vision goggles (NVG) is more likely to experience motion sickness because:

○ a) NVG restrict the field of view

○ b) Colour distortion of the visual field induces nausea

○ c) The weight of the NVG changes the center of rotation of the head, disrupting vestibular signals

○ d) People are more susceptible to motion sickness at night

16. The organs responsible for sensing linear accelerations along the rostral-caudal axis are:

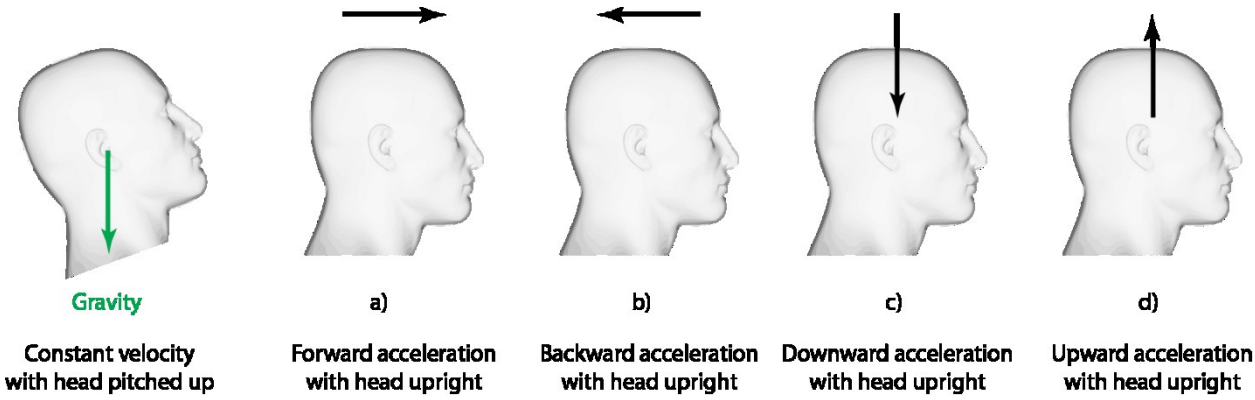
○ a) The saccule only

○ b) The utricle only

○ c) The saccule and utricle

○ d) The semicircular canals only

○ e) The semicircular canals and otolith organs together



17. In the illustration above, the organs responsible for sensing linear acceleration in the rostral-caudal axis will ambiguously interpret movement at a constant velocity and the head pitched up as equal to:

○ a) Forward acceleration with the head upright

○ b) Backward acceleration with the head upright

○ c) Downward acceleration with the head upright

○ d) Upward acceleration with the head upright

18. The implication of the above illustration is that:

○ a) Pilots should take care to change acceleration or pitch independently during flight manoeuvres

○ b) Aircraft designers must keep pilot seat backs vertical to prevent pilot disorientation

○ c) Pilots can perceive that an aircraft is pitching up or down, when in reality it is changing velocity

○ Pilots must only rely on visual information during flight

19. Because of illusions in linear perspective, when a pilot makes an approach over terrain that slopes down toward the runway the pilot is likely to:

○ a) Perceive the approach as too shallow, and unnecessarily make their approach steeper

○ b) Perceive the approach as too steep and unnecessarily make the approach shallower still

○ c) Perceive the approach as too high, and unnecessarily make the approach too low

○ d) Perceive the approach as too low and unnecessarily make the approach too high

20. Presbyopia is a condition affecting:

○ a) Primarily younger aircrew who tend to only notice objects in their near field of vision

○ b) Primarily older aircrew as the eye loses its ability to focus on near objects

○ c) Aircrew of any age experiencing GLOC

○ d) Primarily older aircrew as the eye loses its ability to focus on distant objects

○ e) Aircrew of any age because pressure changes result in IOC changes that causes inability to focus on near objects

21. Pilots and SAR tech must meet the visual testing criteria:

○ a) For the V2 standard for both near and far vision, but far vision will determine the category

○ b) For the V2 standard for both near and far vision, but near vision will determine the category

○ c) For the V2 standard for both near and far vision, but better of the two will determine the category

○ d) For the V2 standard for both near and far vision, but the worse of the two will determine the category

22. Aircrew may receive either LASIK or PRK surgery to correct vision, however:

○ a) Only PRK works on myopia

○ b) They must also wear corrective eyewear for a 3 month period following surgery

○ c) Only LASIK corrects for astigmatism

○ d) Crew members are temporarily grounded following surgery pending re-evaluation

23. Protran and deutran defects in vision are dangerous for pilots because:

○ a) They make it difficult to distinguish the colours and significance of signals and lights

○ b) They make it difficult to focus on runway markings, particularly when tired or stressed

○ c) They make it difficult to focus on near objects, causing problems when reading instruments

○ d) They make it difficult to perform the Titmus Fly test, signifying poor stereoscopic vision

24. The following visual conditions are disqualifying (select 5):

☐ a) IOP greater than 28 mm Hg

☐ b) Hyperopia

☐ c) IOP less than 28 mm Hg

☐ d) Cataracts

☐ e) Retinal tear

☐ f) Vision at V4

☐ g) Corrected vision

☐ h) Iritis

☐ i) Strabismus >/= 10 P.D.

25. Fighter pilots have a higher incidence of cervical spine degeneration than age matched controls most likely due to:

○ a) Helmets, visors, and helmet-mounted kit

○ b) Need to maximize cervical range of motion for best situational awareness

○ c) Increased gravitational forces caused by aeronautical manoeuvres

○ d) Osteoporosis associated with lower gravity environments at altitude

26. Disqualifying musculoskeletal conditions for pilots include (select 4):

☐ a) Active infections such as osteomyelitis or septic arthritis

☐ b) Scoliosis > 20°

☐ c) Mal-union of any fracture

☐ d) Scoliosis > 10°

☐ e) Mal-union of fractures interfering with duties

☐ f) Spondylolysis or Spondylolisthesis

☐ g) Articular Facet Syndrome

27. Wind flail injuries on ejection affect:

○ a) Limbs only

○ b) Thorax (flail chest)

○ c) Limbs and c-spine

○ d) C-spine only

28. Psychiatric disorders that disqualify pilots includes (select 4):

☐ a) Hx of acute stress disorder even if seen as resolved with psychiatric re-evaluation

☐ b) Hx of depressive disorder with any use of antidepressant medication, regardless of subsequent mood stability

☐ c) Hx of attempted suicide

☐ d) Hx of any ADHD

☐ e) Hx of ADHD with Ritalin use

☐ f) Hx of sleep disorder with somnambulance

☐ g) Bipolar Disorder Type 1 or 2

☐ h) Hx of agoraphobia even if seen as resolved with psychiatric re-evaluation

29. Alcohol consumption in aviation is dangerous because:

○ a) Increases G tolerance and decreases hypoxia

○ b) Increases G tolerance and increases hypoxia

○ c) Decreases G tolerance and decreases hypoxia

○ d) Decreases G tolerance and increases hypoxia

30. An aircrew member diagnosed with alcohol abuse or dependence:

○ a) Must automatically be recommended unfit for flying

○ b) Will be recommended unfit for flying only if they do not comply with voluntary treatment

○ c) Must undergo mandatory treatment, regardless of fitness for flying

○ d) Must undergo mandatory treatment, and be recommended unfit for flying

31. After a permanent Air Factor flying category has been assigned, any member who relapses after diagnosis of alcohol abuse or dependence will lose active flying status:

○ a) Immediately

○ b) Only if the relapse is not social or occasional consumption

○ c) Pending a review of the member's history

○ d) Only if the member refuses treatment for the relapse

32. Divers and pilots both experience abnormal environments however:

○ a) ENT exams are more important for pilots because of rapid altitude change

○ b) ENT exams are more important for divers because of rapid pressure change

○ c) ENT exams are more important for pilots because of risk of disorientation

○ d) ENT exams are more important for divers because of risk of disorientation

33. With CFEME approval, a flight surgeon can prescribe Temazepam 15 mg. to facilitate sleep during stop-overs on long-haul transport Ops, to facilitate off-nominal sleep (e.g. daytime sleep) with the crew member returning to duty:

○ a) Under no circumstances whatsoever

○ b) 8 hours after ingestion, and fewer than 5 days of consecutive use

○ c) 24 hours after ingestion, and fewer than 3 days of consecutive use

○ d) At the discretion of the flight surgeon

34. Which is the following gastrointestinal conditions is considered disqualifying for aircrew selection:

○ a) Medication induced gastritis

○ b) Post-traumatic bleeding

○ c) Non-infectious hepatic cyst

○ d) Pancreatic pseudocyst

35. Which of the following sites of metastatic disease if of the greatest potential concern from an aviation perspective?

○ a) Bone

○ b) Brain

○ c) Liver

○ d) Lung

36. A CAF diver had an uncomplicated inguinal hernia repair one year ago. You're asked to do a Type 1 (full) PHA on the diver because there are no Dive MO's on base. The diver tells you that he has an area of decreased sensation extending from the healed incision site to the upper medial thigh. You:

○ a) Recommend "unfit diving" because he has neurologic symptoms that might be mistaken for decompression sickness

○ b) Place the diver on a TCAT for 6 months, to see if neurologic symptoms will resolve

○ c) Send the diver back to his surgeon for an opinion

○ d) Document the area of decreased sensation but don't recommend that he be "unfit diving"

Thank you!

The time you have given to complete this test will help us to improve training and courses in the CAF.

Annex B Student Course Engagement Questionnaire

SCEQ

Please enter your participant code:

The Student Course Engagement Questionnaire

To what extent do the following behaviours, thoughts, and feelings describe you, in this course. Please rate each of them by selecting one item on the following scale: 1 = not at all characteristic of me, 2 = not really characteristic of me, 3 = moderately characteristic of me, 4 = characteristic of me, 5 = very characteristic of me

1. Raising my hand in class.

1 2 3 4 5

○ ○ ○ ○ ○

2. Participating actively in small group discussions.

1 2 3 4 5

○ ○ ○ ○ ○

3. Asking questions when I don't understand the instructor.

1 2 3 4 5

○ ○ ○ ○ ○

4. Doing all the homework problems.

1 2 3 4 5

○ ○ ○ ○ ○

5. Coming to class every day.

1 2 3 4 5

○ ○ ○ ○ ○

6. Contacting the instructor after class to review assignments or tests, or to ask questions.

1 2 3 4 5

○ ○ ○ ○ ○

7. Thinking about the course between lectures.

1 2 3 4 5

○ ○ ○ ○ ○

8. Finding ways to make the course interesting to me.

1 2 3 4 5

○ ○ ○ ○ ○

9. Taking good notes in class.

1 2 3 4 5

○ ○ ○ ○ ○

10. Looking over class notes between classes to make sure I understand the material.

1 2 3 4 5

○ ○ ○ ○ ○

11. Really desiring to learn the material.

1 2 3 4 5

○ ○ ○ ○ ○

12. Being confident that I can learn and do well in the class.

1 2 3 4 5

○ ○ ○ ○ ○

13. Putting forth effort.

1 2 3 4 5

○ ○ ○ ○ ○

14. Being organized.

1 2 3 4 5

○ ○ ○ ○ ○

15. Getting a good grade.

1 2 3 4 5

○ ○ ○ ○ ○

16. Doing well on the tests.

1 2 3 4 5

○ ○ ○ ○ ○

17. Staying up on the readings.

1 2 3 4 5

○ ○ ○ ○ ○

18. Having fun in class.

1 2 3 4 5

○ ○ ○ ○ ○

19. Helping fellow students.

1 2 3 4 5

○ ○ ○ ○ ○

20. Making sure to study on a regular basis.

1 2 3 4 5

○ ○ ○ ○ ○

21. Finding ways to make the course material relevant to my life.

1 2 3 4 5

○ ○ ○ ○ ○

22. Applying course material to my life.

1 2 3 4 5

○ ○ ○ ○ ○

23. Listening carefully in class.

1	2	3	4	5
○	○	○	○	○

List of Symbols/abbreviations/acronyms/initialisms

BAvMed	Basic Aviation Medicine
CAF	Canadian Armed Forces
CFEME	Canadian Forces Environmental Medical Establishment
CFITES	Canadian Forces Individual Training and Education System
DRDC	Defence Research and Development Canada
IT&E	Individual Training and Education
MSLQ	Motivated Strategies for Learning Questionnaire
PO	Performance Objective
QS	Qualification Standard
SCEQ	Student Course Engagement Questionnaire
SOM	School of Operational Medicine
TP	Training Plan

DOCUMENT CONTROL DATA

*Security markings for the title, authors, abstract and keywords must be entered when the document is sensitive

1. ORIGINATOR (Name and address of the organization preparing the document. A DRDC Centre sponsoring a contractor's report, or tasking agency, is entered in Section 8.) DRDC – Toronto Research Centre Defence Research and Development Canada 1133 Sheppard Avenue West Toronto, Ontario M3K 2C9 Canada	2a. SECURITY MARKING (Overall security marking of the document including special supplemental markings if applicable.) CAN UNCLASSIFIED 2b. CONTROLLED GOODS NON-CONTROLLED GOODS DMC A

3. TITLE (The document title and sub-title as indicated on the title page.)

Learning Retention Effects and Student Engagement for a Basic Aviation Medicine Course: A Pilot Study of Baseline Performance

4. AUTHORS (Last name, followed by initials – ranks, titles, etc., not to be used)

Martin, B. C. W.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total pages, including Annexes, excluding DCD, covering and verso pages.)	6b. NO. OF REFS (Total references cited.)
April 2020	37	18

7. DOCUMENT CATEGORY (e.g., Scientific Report, Contract Report, Scientific Letter.)

Scientific Report

8. SPONSORING CENTRE (The name and address of the department project office or laboratory sponsoring the research and development.)

DRDC – Toronto Research Centre
Defence Research and Development Canada
1133 Sheppard Avenue West
Toronto, Ontario M3K 2C9
Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
04g - Personnel Generation	
10a. DRDC PUBLICATION NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	**10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)**
DRDC-RDDC-2020-R041	

11a. FUTURE DISTRIBUTION WITHIN CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

Public release

11b. FUTURE DISTRIBUTION OUTSIDE CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

12. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Use semi-colon as a delimiter.)

Competency-Based Training; Learning and Training; Testing; Qualification Standard; Retention; Instructional Practice; Aviation Medicine

13. ABSTRACT (When available in the document, the French version of the abstract must be included here.)

Classroom technologies such as SmartBoards, collaboration software and hardware and classroom response systems are seen as enablers in the effort to modernize individual training and education (IT&E). It is impossible to verify and validate the efficacy of such tools without first establishing baseline measures of student performance in courses, and engagement with the learning experience. Most tests in the Canadian Forces are pass/fail criterion-referenced, and do not provide sufficiently granular data to understand the influence of specific interventions. Using norm-referenced measurement, we tested students on how much they knew about aviation medicine before and after the Basic Aviation Medicine (BAvMed) course offered through the School of Operational Medicine (SOM) within the Canadian Forces Environmental Medicine Establishment (CFEME). Students performed a test that paralleled the content and structure of the qualification test at one week before, immediately after, six weeks after, and six months after the course. Test scores improved by 13 more correctly answered questions from before to after the course, and we observed very little decay after six months. We also compared test scores from the parallel and qualification tests and found there was no difference between the two, suggesting that the parallel test is a reasonable measure of student achievement for this course. Last, a questionnaire on student engagement revealed no correlation between test scores and any measure of engagement. These results provide performance benchmarks and methodological validation for future investigations of educational technology in the context of the BAvMed course.

Les technologies en classe, comme le tableau numérique interactif ou smartboard, le matériel et les logiciels de groupe, ainsi que les systèmes d'intervention pédagogique sont perçus comme des outils dans les efforts de modernisation de l'instruction individuelle et l'éducation (II et E). Il est impossible de vérifier et de valider l'efficacité de ces outils sans avoir d'abord établi une base de référence pour évaluer le rendement des étudiants en classe et leur engagement à l'égard de l'expérience d'apprentissage. La plupart des tests dans les Forces armées canadiennes fonctionnent selon le critère réussite ou échec et ne fournissent donc pas de données suffisamment détaillées pour nous permettre de comprendre l'influence d'une intervention en particulier. Au moyen de mesures normatives, nous avons évalué les connaissances des étudiants en matière de médecine aéronautique avant et après avoir suivi le cours élémentaire de médecine aéronautique offert par le Centre de médecine environnementale des Forces canadiennes (CMEFC). Les étudiants se sont soumis à un test dont le contenu et la structure sont analogues à ceux du test de qualification une semaine, immédiatement après, six semaines après et six mois après le cours. Les étudiants ont répondu correctement à 13 questions de plus après le cours, comparativement au premier test, et nous avons constaté très peu d'érosion des connaissances après six mois. Nous avons également comparé les résultats du test analogue au test de qualification et nous n'avons trouvé aucune différence entre les deux, ce qui suggère que le test analogue est une mesure raisonnable du rendement des étudiants de ce cours. Finalement, un questionnaire sur l'engagement des étudiants n'a révélé aucune corrélation entre les résultats aux tests et leur degré d'engagement. Ces résultats fournissent des seuils de référence pour le rendement et une validation méthodologique pour de futures enquêtes sur les technologies éducatives dans le cadre du cours élémentaire de médecine aéronautique.