Remarks on Assessment in Cadets Common Technology Courses in NDU

Antti J. Rissanen and Kalle O. Saastamoinen

Abstract—**The National Defence University (NDU) trains officers for the Finnish Defence Forces (FDF) and the Border Guard. This paper presents NDU's practices for assessment in the common part of technology education in cadet training. The starting level of education at NDU involves earning a Bachelor's degree in military science. Methods for collecting data on knowledge and skills are reviewed to understand learning measurement. The evaluation of skills is an integral part of assessment. For the evaluation of knowledge, different cases are selected to consider practical aspects of assessment.**

Index Terms—**Assessment, knowledge, performance test, skills, summative test.**

I. INTRODUCTION

The National Defence University (NDU) is a training institution responsible for educating the future leaders of Finland's armed forces. Officer cadets gain tertiary qualifications and the fundamental skills and knowledge needed to undertake commissioned officer roles in the Army, Navy, Air Force, and Frontier Guard. The three-year bachelor courses cover military strategy and tactics, military technologies, and military history. Academic studies are combined with leadership, communication, and military training. In addition, for further studies are offered by the university [1].

NDU provides education on the officer's profession according to the Bologna process [2]. Military technology is the major learning focus at NDU. With basic scientific education and selected technology-related courses, technology is an essential part of the general curriculum in cadets' basic studies. The learning aims may be expressed in three equally important parts, which should not be separated: 1) gaining basic scientific knowledge, 2) understanding how technology utilizes scientific results, and 3) gaining the ability to manage practical exercises within the working environment. Measuring skills and controlling learning with grading to provide feedback on each cadet's individual learning progress is relatively challenging and time-consuming.

Course feedback and other new practices that estimate educational processes are both interesting and challenging for active teachers, but they have drawbacks (e.g., exams requiring excessive administration take resources away from instruction or other efforts to improve educational outcomes.

II. ASSESSMENT GENERALLY

Broadly, course-related assessment involves two different types of activities. Thus, assessment tools aim to first collect data on the knowledge and skills that students have learned (measurement) and then to estimate each student's level of learning (evaluation) [3]. In most university-level courses, the measurement and evaluation aspects of assessment can be done in many ways. For example, to determine how much learning has occurred, students may be asked to take exams, respond to oral questions, complete homework exercises, take presentations, or enter a virtual reality simulator [4].

At NDU, assessments are aimed at ranking students or simply providing general accountability. In common Science, Technology, Engineering, and Mathematics (STEM) studies, students take exams at the end of each course period. Hence, summative assessment is dominant, even though student learning with self-assessment is supported. It has been claimed that typical exams are based on tradition or simply related to the material from the textbook publisher. However, the curricula at NDU define most of the courses in such a way that no suitable external ready-made assessment material is available. A well-prepared assessment reflects those concepts and skills that the lecturer emphasized either in lectures or in the learning management systems (LMS). Teachers and students share responsibility for learning. It is recognized that an instructor's effectiveness is not defined on the basis of what he or she does as a teacher but rather on what his or her students are able to do.

In a military context, all technology-related courses include multiple learning objectives, some of which can be considered as tacit knowledge, and as such they are largely unmeasurable. Generally, exams should measure student's knowledge and measureable skills [5]. However, exams could also support other educational aspects such as student learning in the long run, creative thinking, and motivation.

Careful examination analysis offers information to teachers on how to develop courses. At NDU, lecturers have attempted to develop applied science courses such that the related exams and other assessment activities have versatile functions for learning. The evaluation of professional skills in the officer's profession involves multiple challenges where accountability is considered.

III. TEACHING FOR LEARNING

A. Overloading Issues

Continuous development in science and technology are often claimed to produce a knowledge explosion. Likewise, demands to improve people's personal ability to learn scientific knowledge, to obtain new information, and to resolve real-world problems are widely expressed. It is not enough that students receive professional knowledge; they must have the ability to apply knowledge and skills in totally new situations and unfamiliar environments.

Cadets' formal level of knowledge and personal interests varies. In general courses, the requirements are set appropriately for non-physics or non-mathematics majors. According to the curricula and the semester plan, only compact and tightly scheduled science and technology courses are feasible.

A sustainable approach in any technology-rich course for cadets is to teach concepts and principles that are applicable to any device and will be relevant to the future profession. This is done mostly on a theoretical basis. A relatively limited lecturing frame must be utilized, selecting only essential topics for lectures and controlling the amount of materials and homework that are presented in the LMS. Most homework is done in small groups led by peers, so time for teamwork must also be planned. These facts limit the depth of learning goals and mean that assessment must follow the annually appointed learning resources.

Following this approach, we promote a constructivistic student-centered methodology that helps students to develop their skills in the discipline as well as their capabilities as team members. Such skills are useful to learn, not only to pass the course but also for the future profession. New technologies open up opportunities for new ways to achieve education aims. On the other hand, when looking at how examinations are held, there are many practical reasons for using paper-and-pen types of exams, whereas the assessment structure itself might be ready for application in the LMS.

IV. Research Problem

As an accredited university-level institute, NDU must measure student learning. The measurement technique or assessment protocol needs to be suitable for its purpose, and the modern university establishment is continuously challenged to do things better. One of the key issues is improving general efficiency by looking at applied processes and resource utilization. Exam practices and all other student evaluation procedures must be subject to quality inspections and plans for improvement. Hence, tools like self-evaluations performed by teachers with gathered course data may open up opportunities for better outcomes, providing tools for minimizing any detected disadvantages and improving the way in which teacher resources are allocated. In the technology courses included in the general curriculum, we have classified learning goals into two categories:

1) Officer's professional skills and university-level study skills;
2) STEM-related scientific knowledge and course-related applied knowledge details.

The major question is how can we effectively assess these two different goals?

1) Can we make any observations on skills using less grade-oriented evaluations?
2) Can a paper-and-pencil variant of an electric form-based exam be effective enough for evaluation and still be relatively easy to grade?

To find a feasible solution, we present selected tools and examine the way in which the evaluation practice is defined in the literature. Our solution is presented using a case study approach examining second-year courses. The cases contain skill- (performance test) and knowledge-type evaluation approaches.

V. Methodology

Action research supports educators in seeking out ways in which they may systematize their problem setting and take further steps towards solutions, thus enhancing the quality of education. Action research is designed to bridge the gap between research and practice. Kemmis and McTaggart [6] propose that action research forms a spiral of process stages in which the process may not be as neat as the spiral of self-contained cycles of planning, acting, observing, and reflecting suggests. As a good example of this roughness (case number 3: combination test), one course has been studied multiple times, and multiple interventions have been made to its environment in recent years. Moreover all of the courses under study have roots in earlier curricula as well as links to other courses for cadets and master level courses at NDU.

Therefore, the emphasis of this study follows O'Leary's model [7]: action research is considered as an experimental learning approach, with aims including the continuous need to refine the methods, data, and interpretation. Action research methodology describes a practical approach to solving a detected learning environment problem. In this study, the examination and performance tests are studied to provide better knowledge of learning outcomes. These two items are related to the structure of how teaching is applied. The cycle nature in development is presented, but a detailed presentation is avoided.

Courses at NDU are relatively intensive and the scheduling of the grading or assessment is generally discussed at the beginning of each course. Hence, assessment appears in many sections of teaching entities.

VI. Assessment Tools

A. Facts and Ways to Measure Student Learning

In most of university classes, student learning is estimated frequently. Suitable tools for assessment may vary according to the discipline, course level, and amount of participants in the course. Instructors have several ways to measure what students have learned. The objectives in measurement can be classified in terms of two categories: knowing about something and knowing how to do something.

As an evaluation, written tests attempt to assess the scale and accuracy of a student's knowledge. Other measures that attempt to assess how well students can do something are referred to as performance tests. Both types of evaluations are handy in any skillful teacher's assessment toolbox. Which type is used, and to what extent, will depend on the purpose

behind the assessment. The amount of time instructors spend assessing student learning depends on the number of students, the discipline, grading demands, and whether other personnel are able to participate in the process. Nevertheless, most teachers spend a substantial portion of their working hours assessing student learning, and much of this assessment activity involves planning, streamlining, and scoring practices.

B. Measurement

Measurement means the assignment of numbers to certain attributes in accordance with explicitly expressed rules [8]. As examples: in shooting tests, the number of hits inside the correct area represents the score or the student's level; in diagnostic tests of mathematical skills, scores are determined by counting the number of tasks that have been solved correctly [9].

C. Formative Evaluation

A second reason for assessing students is to monitor their progress. The main things that teachers want to evaluate periodically are whether students are keeping up with the pace of instruction and if they understand all the material that has been covered. For students whose pace of learning is slower than average or for those whose understanding of learning topics is problematic, we can offer supplementary or remedial instruction in the LMS. This type of assessment is called formative evaluation, and it is not aimed at grading but rather at supporting learning.

Finally, the assessment of student performance may have positive effects on various aspects of learning and instruction. Knowing principles of assessment may cause students to ponder what is important to learn and affect their motivation and self-perceptions of competence, structure their approaches to and timing of personal study, and affect the development of their learning strategies and skills [10].

D. Summative Evaluation

The leading aim of most assessments is to provide to all parties of an educational institution a clear, meaningful, and useful summary or accounting of how well their students have met the course's objectives. For this purpose, the most commonly utilized assessment tool is the summative test, according to which the instructor in charge of a course determines the grade of each student based on exam success. In Finland, students enroll in courses on either a numeric grade or a Passed/Not Passed basis. The final course exam is called a summative evaluation since its primary purpose is to sum up how well a student has performed. Summative assessment has been widely studied, and it has been observed to give data not only on learning but also on students' motivation for learning [11].

E. Written Tests

Most written tests are comprised of a mixture of one or more of the following item types: selected response (e.g., multiple choice, true–false, and matching), short answer, and essay. Each item type is aimed at measuring how much a pupil knows about a particular subject.

1) Selected-response test

In a selected-response test, the student reads a brief statement and selects one of the provided alternatives as the correct answer. Selected-response tests are typically made up of multiple-choice, true–false, or matching items. This type of test is often viewed as an "objective" test because it has a direct fact-based scoring system.

Selected-response tests are typically used when the primary goal is to assess what might be called foundational knowledge. This is the basic information that students need in order to perform more demanding tasks in the subject area.

The most obvious advantage of a selected-response test is its efficiency. The instructor can ask many questions, and the student may answer them quickly. For the instructor, the major advantage is ease and reliability of scoring. With the aid of a scoring template or as a tool in the LMS, many tests can be quickly made, tested, and scored.

On the other hand, it might be easiest to formulate items that represent the lowest level of applied learning taxonomy (e.g., verbatim knowledge in Bloom's taxonomy). As a result of seeing example tests, students tend to focus on simple memorization rather than on meaningful learning. An even bigger disadvantage might be that there is no measure of what students can do with the knowledge [3].

2) Short-answer response test

With a short-answer response test, after a question is presented, the student is advised to provide a relatively brief answer. The optimal response may consist of a phrase, explanation of symbols, or even a few sentences. Wyse short-answer tests are designed so that they can be scored quickly, accurately, and consistently. For the instructor, the major advantage of these tools is the fact that it is relative easy to find and formulate such tasks from the teaching material. Moreover, assessment can include tasks that measure both broad and in-depth knowledge. Unfortunately, when short answers are requested, students tend to limit their processing. Therefore, the given answers might not give an indication of the depth of learning. Sometimes unexpected but plausible answers might stop the grading process and require upgrades to the preliminary planned score model.

3) Essay test

In essay tests, the student receives a general directive or some guiding words to explain the theme and ideas for applying it. The following example gives an idea of an essay question in NDU's military context: "Describe passive and active sensor system principles, give examples and look for the meaning for surveillance and target acquisition applications."

Most essay tests reveal how well students can remember, organize, and communicate learned information. When the course's demands are known, students are more likely to try to learn the course material in a constructive way. Therefore, when the essay test procedure is practiced and the test itself is formulated properly, the answers may reflect higher-level abilities such as analysis, synthesis, and evaluation.

Due to the contents and variations in expressions as well as in details, such essay answers are often time-consuming to grade. Moreover, variations in approaches to answering may produce serious fluctuations in loosely formulated essay-writing tasks. Therefore, instructors' demands and models for desired answer structure should be emphasized

and, if possible, practiced. [3]

F. Performance Test and Test types

Teachers must sometimes create a tool that differs considerably from basic essay writing, even if it often includes text-based items (e.g., writing plans or documenting). Performance tests attempt to assess how well students use foundational knowledge to perform complex tasks under more or less realistic conditions [12]. At the low end of the realism spectrum, students may be asked to construct a map, interpret a graph, or write an essay under highly standardized conditions. That is, everyone completes the same task in the same amount of time and under the same conditions. At the high end of the spectrum, students may be asked to conduct a science experiment or write an essay regarding a technical artifact or how to apply it [3].

1) Direct assessments

Here, a student is asked to provide either an oral, or a written answer to a practical task. In a larger study group, a direct writing assessment tests are applied in a way that students are asked to write about a specific topic under a standard set of conditions.

2) Demonstrations

In courses related to science and technology, students can be asked to give demonstrations. In such demonstrations, they are required to show how well they can use previously learned knowledge or skills to solve a somewhat unique problem. Due to limits in resources, such tasks are usually simple. Examples might be presenting a group report for the class or repeating a short part of a laboratory task.

3) Exhibitions

Due to the many choices offered by modern technology, it is sometimes difficult to classify practice-oriented performance test results as demonstrations or exhibitions. When students are showing constructs such as drawings, photographs, video clips, and even PowerPoint presentations, the assessment is categorized as an exhibition.

VII. Assessment Practice at NDU

A. Ways to Evaluate Skills and Knowledge Application

The assessment of skills in a larger learning group is both time-consuming and resource-intensive. Formative non-grading assessment during lectures allows tuning teaching, but it is not necessarily linked to grading itself. Discussion about the learning targets and assessment procedures helps cadets in the long run in their exam preparations.

Due to personal differences in STEM-related skills among cadets, only growth potential is measured in technology-oriented courses. Therefore, skill-related learning goals are not graded but only observed, and if the performance is acceptable, then brief feedback is given. If the cadet fails in the performance test, then he or she is asked to revise the failed items that have been identified in the feedback. Due to the fact that every course at NDU is required to support officers' professional skill development, some type of skill-related evaluation must be included for good assessment practice.

1) Case: Performance evaluation

In the first quarter of most technology-related courses, relatively simple or straightforward homework is given. The idea of the task is presented briefly in lectures, and part of the material or only the task text is given in the LMS. After a few days, the tasks are presented in lectures with an opportunity for discussion. After the discussion, more guidance is given, and on the final day the exercise must be returned to the portal.

The purpose is to produce tools for making more effective calculations or comparisons. More broadly, the goal is to learn some of the basic skills needed to utilize many features of Excel. After the supporting sessions, students load their solutions into the portal. The instructor estimates the results and provides feedback. If any student's answer fails or is inadequate, then the hidden additional self-learning course material is opened in the portal and he or she is recommended to complete at least the first two sections of the given learning package.

If the general course schedule allows, then a few new homework assignments with assessments may be given before the final homework. At the end of the course, study groups are asked to prepare a short presentation about one of the topics in which background and theory were presented in lectures.

The presentation approach may vary from a basic PowerPoint presentation to an experimentation video. This performance-type work receives a passing grade when the theme, presentation, report in the portal, and final presentation are completed successfully. The work effort from team to team and also inside each study group differs.

B. Our Grading Exam on Knowledge

A successful NDU exam attempts to measure more than substantive knowledge; cadets' ability to apply essential information in the subject area as well as their critical thinking, abstract reasoning, and innovation capabilities are considered. Following this approach, exams are demanding at first sight, even though the structure, principles, and preliminary training are given in lectures and in our LMS, which is a standard Moodle-based learning management system.

1) Case: Selected-response test in moodle

In one of our system technology courses, the summative examination was based fully on Moodle. Videos and selected text material were presented so that selections in the multiple-choice test measured both observation skills and memorization. As such, the examination preparation was demanding, and the future utilization was left open. Some issues, however, still need better consideration and planning.

2) Case: Classical selected-response test

In the communication technology, much effort was allocated to student group works and to getting visiting professionals to present how communication technologies and protocols are utilized in the Finnish Defence Forces. The connected exam was a straightforward, multiple-choice-based, quickly gradable paper-and-pencil exam. The cadets indicated that they liked the general

arrangement, but the best students felt that they were unable to show their full competence in the exam. This type of exam has limits with regard to aspects emphasizing creativity.

3) Case: Combination test

In the Surveillance and Target Acquisition technology course, both technical issues and atmospheric phenomena are learned and applied to military purposes. The summative exam starts with two short answer questions about technology. These are followed by 12 true–false statements. Next, the cadets are asked to explain a graph presenting three different thermal radiation curves followed by a second graph and a question: name the main components in the [atmospheric attenuation] graph. The last task is composed of two choices, either calculus on radar principles or an essay on thermal imaging in military contexts. Equations for the questions and some more aids are provided. Cadets generally find this type of examination demanding, but it is effective for measuring knowledge in major course topics.

For grading, only the essay requires extra effort on the part of teachers. In the end, this type of an exam provided teachers with the most information regarding how to develop and streamline the course content.

C. Assessment of Knowledge and Grading

STEM subjects are commonly related to the academic side of teaching, and they consist of many knowledge types and levels. At NDU, criterion-referenced grading is utilized. Hence, we assign numeric grades on the basis of the percentage of test items answered correctly. A grade of 5 (excellent knowledge/skills) is awarded to anyone who correctly answers at least 95 percent of a set of test questions. To pass a course, about 45% of the available raw points must be gained. The grades 4, 3, and 2, and sometimes even the middle halves, are scaled linearly between the end points. This type of grading system is considered fair, as cadets are provided with realistic and specified criterion levels. If a full exam is not included in the academic curricula or if the assessment is part of a larger study set, then a Passed/Not Passed grade is given according to presentations or written material. However, new ideas are needed to produce different types of items for both LMS-based electronic tests and paper-and-pencil tests.

VIII. CONCLUSION

Assessment of a larger learning group is both time-consuming and resource intensive. From our experience, we conclude that a more versatile exam structure might motivate and inspire students. In the long run, an informative exam with a profession-related structure could lead to better student learning and provide satisfaction to the teaching team. However, several exam types are needed in our practice. The course content will dictate which exam type is the most practical selection.

In any Moodle configuration, the following tests are easily applied:
- Short answer
- True–False
- Explain some given phenomena
- Explain data set or graph

- Essay on selected topic

On the other hand a full calculation procedure requires specific add-on tools and practice to effectively measure learning in the Moodle LMS. Even though findings regarding physics exercises in the Moodle environment have been long available [13], without proper add-ins we are forced to continue with paper-and-pencil examinations if we want to grade and support calculus practice.

Our findings suggest that by following a performance test approach, some of the key issues in young officers' career skills could be tested and strengthened, even in relatively theoretical and technology-rich courses. This means that evaluations employing less numeric grading evaluation might be fruitful if students are supported and if the preparations are well designed. Skill and knowledge evaluation are not separate areas. Performance-type tests can also offer information about how well learned knowledge is understood and applied. We also noticed that with the professional skill evaluation (performance test), student motivation improved during the last lecturing session. Further, development practices in the Moodle environment will make the future workload of teachers and tutors more manageable.

REFERENCES

[1] NDU. [Online]. Available: http://maanpuolustuskorkeakoulu.fi/en/studies

[2] B. M. Khemu and U. Teichler, "Which direction for bachelor and master programmes? A stocktaking of the Bologna process," *Tertiary Education and Management*, vol. 12, no. 4, pp. 269-282, 2006.

[3] J. Snowman, R. McCown, and R. Biehler, *Psychology Applied to Teaching*, 12th ed. Boston, MA: Houghton Mifflin, 2009, ch. 14, pp. 475-512.

[4] R. Stiggins, J. Arter, J. Chappuis, and S. Chappuis, *Classroom Assessment for Student Learning: Doing it Right - Using it well, Portland*, OR: Assessment Training Institute, 2009, ch. 1, pp. 1-28.

[5] J. Biggs, "What do inventories of students' learning processes really measure? A theoretical review and clarification," *British Journal of Educational Psychology*, vol. 3, no. 1, pp. 3-19, 1993.

[6] S. Kemmis and R. McTaggart, "Participatory action research," in *Handbook of Qualitative Research*, 2nd ed., N. Denzin and Y. Lincoln, Eds. Thousand Oaks, CA: Sage, 2000, pp. 567–605.

[7] Z. O'Leary, *The Essential Guide to Doing Research*, London: Sage, 2004.

[8] M. J. Gierl, "Making diagnostic inferences about cognitive attributes using the rule-space model and attribute hierarchy method," *Journal of Educational Measurement*, vol. 44, no. 4, pp. 325–340, Winter 2007.

[9] A. Rupp, J. Templin, and R. Henson, *Diagnostic Measurement: Theory, Methods, and Applications*, New York, NY: Guilford Press, 2010, ch. 1-3.

[10] T. J. Crooks, "The impact of classroom evaluation practices on students," *Review of Educational Research*, vol. 8, no. 4, pp. 438-481, 1988.

[11] W. Harlen and R. Deakin, "A systematic review of the impact of summative assessment and tests on students' motivation for learning" *Research Evidence in Education Library*, London: EPPI-Centre, 2002, Issue 1.

[12] G. E. Miller, "The assessment of clinical skills/competence/ performance," *Academic Medicine*, vol. 65, no. 9, pp. S63-67, 1990.

[13] T. Martín-Blas, and A. Serrano-Fernández, "The role of new technologies in the learning process: Moodle as a teaching tool in Physics," *Computers & Education*, vol. 52, no.1, pp. 35-44, 2009.

Antti J. Rissanen was born in Helsinki. He obtained a licentiate in technology degree (electric engineering) in 1992 from the University of Technology in Tampere and his doctoral degree in physics in 2011 from the University of Helsinki in Finland.

In 1997, he joined the National Defence University (Finland) as a senior lecturer in general technology. His research interests include science education, e-learning, and new methods in PER.

He has had various STEM-related teaching experiences, including working as a lecturer in Saimaa University of Applied Science and as a teaching researcher in Lappeenranta University of Technology (LUT).

He has been active in teams developing new methods and technologies for teaching engineering mathematics, for example, in a national project featuring a collaborative network of nine Finnish universities, with Helsinki University of Technology being the lead partner.

Kalle O. Saastamoinen is currently working as a senior lecturer in the National Defence University, Helsinki. He earned his Ph.D. in applied mathematics in 2008 from Lappeenranta University of Technology.

His research interests are computational logic and its applications, AI based expert systems, data classification, military technology research, and teaching technology and its usage.