UiT The Arctic University of Norway

Faculty of Health Sciences

From student to professional – mastering the necessary non-technical skills

Assessment tools and effects of multi-professional simulation training

Katrine Prydz

A dissertation for the degree of Philosophiae Doctor – January 2024

Table of Contents

List of Tables

List of Figures

Acknowledgments

Doing this PhD has been very interesting, really inspiring, quite rewarding, sometimes lonely and occasionally overwhelming. Especially doing research on multi-professional teams during a pandemic was a challenge. At times I did not think I would ever finish, but somehow, I did. I am deeply grateful for all those who have supported me throughout these six and a half years.

My main supervisor Torben Wisborg. You deserve a huge thank you for your invaluable support, everlasting patience, endless knowledge, and strong encouragement throughout all these years, and all these obstacles. I think there were days you thought I would never finish as well. Thanks to both my co-supervisors Peter Dieckmann and David Musson for your invaluable contributions due to your high expertise in the field of non-technical skills. Thanks to Hans Fagertun for your wise insights and explanations when helping me with the statistics.

Thanks to Ingrid Petrikke Olsen who encouraged me to apply for this PhD position. Thanks to the Northern Norway Regional Health Authority (Ref: HNF1368-17) who funded this project. Thanks to Mette Kjær for being a skilled leader that gave med the autonomy, flexibility, and adaptions I needed to survive as an employee in the hospital trusts' new public management system. Also, thanks to my main employer Hammerfest municipality. Despite a rough start, it turned out well. I appreciate the flexibility and support you provided for me to complete this degree.

Thanks to all the students and facilitators that participated in this project, who let me video record them during their training, to all who participated in in the focus group interviews, and to all the raters that contributed with their time and expertise. I really appreciate it.

To all my colleagues, friends and family who have had no clue of what I have been doing, but still have been utmost supportive. Thank you for always believing in me.

Tiril, Henrik and Filip. You are my greatest inspiration and the most important mentors in my life. You have ensured that this PhD was never my primary project. I am truly grateful for that.

Scientific environment

The research in this thesis was conducted from 2017 to 2023. Throughout this period, I served as a PhD student at the Department of Clinical Medicine, UiT, The Arctic University of Norway (UiT), concurrently being an employee at the Finnmark Hospital Trust. I have been a member of the research network of the Finnmark Hospital Trust and the research group Interest at UiT.

My primary supervisor, Professor Torben Wisborg (UiT, Norway), initiated this research project. My co-supervisors were Professor Peter Dieckmann (University of Stavanger, Norway/CAMES, Copenhagen, Denmark) and Assistant Clinical Professor David Musson (McMaster University, Hamilton, Canada).

UiT The Arctic University of Norway

FINNMARKSSYKEHUSET
FINNMÁRKKU BUOHCCIVIESSU

Summary

Newly graduated doctors have a lot of responsibility. The competence they are expected to have consists of medical knowledge, technical skills, but also non-technical skills (NTS). NTS are skills in, communication, decision making, situational awareness and the ability to collaborate. Students should learn this during their years in medical school.

In this thesis, we have taken a closer look at Norwegian medical students' NTS. There were no existing tools to assess Norwegian medical students' NTS, so in the first study we created such a tool. We conducted focus group interviews with experienced doctors, paramedics, user committees and nurses in emergency departments. In the interviews, we mapped which NTS they thought were necessary. We then looked at existing tools and literature and created the NorMS-NTS tool. A tool for assessing Norwegian medical students' NTS.

In the second study, we wanted to validate the tool for use in real life. We wanted the tool to be useful for busy clinicians teaching medical students. We therefore had three experienced healthcare professionals (raters) use the NorMS-NTS to assess medical students' NTS in 20 video recordings with minimal rater training. We then assessed the tool's *interrater reliability, internal consistency,* and *observability* using statistics. The raters also completed a questionnaire about the tool's usability. We found that the tool was usable and had sufficient *interrater reliability, internal consistency* and *observability* for use by novice raters.

In the third study, we compared the NTS performance between students at three different training sites. The three training sites Bodø (urban, decentralized), Tromsø (urban, main campus) and Finnmark (rural, decentralized), were all affiliated with the same university. They had the same exam and learning objectives, but different learning arenas and some different learning activities. The three sites were compared by eight raters viewing a total of 24 videos of students and rating their NTS using the NorMS-NTS. There were eight students from each location and the raters were blinded to the site of training. We performed statistical analyses to determine whether we could observe any difference between training sites. Finnmark had significantly higher level of NTS than Bodø and Tromsø. This finding suggests that high levels of NTS performance can be achieved through rural, decentralized medical education.

Sammendrag

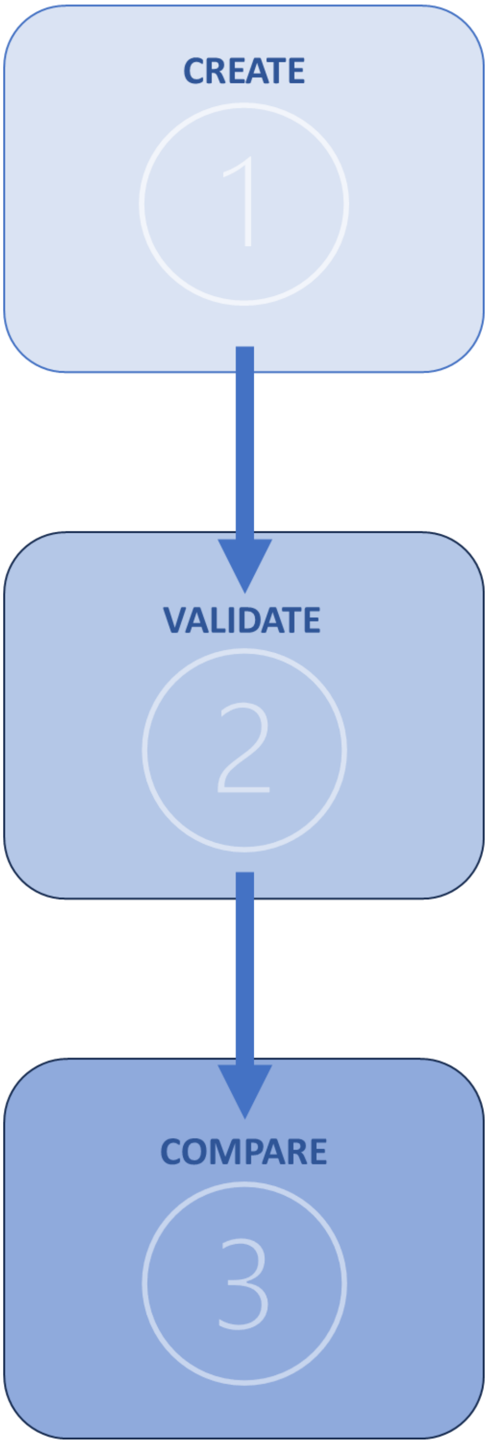
Nyutdannede leger har tidlig mye ansvar. Den kompetansen de forventes å ha består av medisinsk fagkunnskap, tekniske ferdigheter men også ikke-tekniske ferdigheter (NTS). NTS er ferdigheter kommunikasjon, beslutningstaking, situasjonsbevissthet og evne til å samarbeide. Dette bør studentene lære i løpet av årene som medisinstudent.

I denne avhandlingen har vi sett nærmere på norske medisinstudenters NTS. Det var ingen eksisterende verktøy for å vurdere norske medisinstudenters NTS, så i første studie laget vi dette verktøyet. Vi gjennomførte fokusgruppeintervju med erfarne leger, ambulansepersonell, brukerutvalg og sykepleiere på legevakt og i akuttmottak. I intervjuene kartla vi hvilke NTS de mente var nødvendige. Deretter så vi på eksisterende verktøy og litteratur og lagde verktøyet NorMS-NTS. Et verktøy for å vurdere norske medisinstudenters NTS ved observasjon.

I den andre studien ønsket vi å validere verktøyet for bruk i det virkelige liv. Det skulle være brukbart for den travle kliniker som underviser medisinstudenter. Vi fikk derfor tre erfarne helsepersonell (ratere) til å benytte NorMS-NTS til å vurdere medisinstudenters NTS i 20 filmer. De fikk 30 minutter opplæring og en skriftlig bruksanvisning. Deretter vurderte vi verktøyets *interrater reliabilitet, internal consistency* og *observabilitet* ved statistiske beregninger. Raterne fylte også ut et spørreskjema om verktøyets brukervennlighet. Vi fant at verktøyet var brukervennlig og hadde tilstrekkelig *interrater reliabilitet, internal consistency* og *observabilitet* for å kunne brukes av ratere med lite opplæring i bruk av verktøyet.

I tredje studie sammenlignet vi nivå av NTS hos studenter på tre ulike studiesteder. De tre studiestedene Bodø, Tromsø og Finnmark var alle tilknyttet samme universitet, de hadde samme eksamen og læringsmål, men ulik læringsarena og noe ulike læringsaktiviteter. Tromsø var hovedcampus, Bodø var urban, desentralisert og Finnmark rural, desentralisert. De tre stedene ble sammenlignet ved at åtte ratere så totalt 24 videofilmer av studenter og vurderte deres NTS ved bruk av NorMS-NTS. Det var åtte studenter fra hvert sted og raterne var blindet for studiested. Vi gjorde statistiske analyser for å se om vi kunne se noen forskjell på studiested. Finnmark hadde signifikant høyere nivå av NTS enn Bodø og Tromsø. Dette tyder på at man kan oppnå høye nivå av NTS med rural, desentralisert undervisning.

Thesis at a glance



Paper 1:
Create a tool to assess Norwegian Medical Students Non-Technical skills - **The NorMS-NTS tool**

Paper 2:
Collecting evidence of validity for **the NorMS-NTS tool**

Paper 3:
Use **the NorMS-NTS tool** to compare NTS performance at different study sites

List of papers

Paper 1. Prydz K., Dieckmann P., Musson D. & Wisborg T. (2022). The development of a tool to assess medical students' non-technical skills - the Norwegian Medical Students' Non-Technical Skills (NorMS-NTS). *Medical Teacher 2022:1-8.*

Paper 2. Prydz K., Dieckmann P., Fagertun H., Musson D. & Wisborg T. (2023). Collecting evidence of validity for an assessment tool for Norwegian medical students' non-technical skills (NorMS-NTS): usability and reliability when used by novice raters. *BMC Medical Education 2023;23(1):865*

Paper 3. Prydz K., Dieckmann P., Fagertun H., Musson D. & Wisborg T. (Submitted). Non-technical skills of Norwegian medical students at different training sites: A comparative, observational cohort study.

Abbreviations

ANTS Anesthetists' Non-Technical Skills

CBME Competency-Based Medical Education

CRM Crew Resource Management

DTP Decentralized Training Platforms

GP General Practitioner

HFE Human Factors and Ergonomics

ICC Intraclass Correlation Coefficient

NorMS-NTS Norwegian Medical Students' Non-Technical Skills

NPE The Norwegian System of Patient Injury Compensation

NTS Non-Technical Skills

OSCAR Observational Skill-based Clinical Assessment tool for Resuscitation

NOTECHS The Oxford Non-Technical Skills tool

TSD Services for Sensitive Data

UiT UiT The Arctic University of Norway

WHO World Health Organization

1 Introduction

Primum non nocere – first, do no harm (1).

The ancient oath that dedicated physicians follow from the beginning of medical school throughout their professional lives is not only accurate and simple but also challenging. Most physicians make mistakes during their careers (2, 3), which is an inevitable part of their professional lives. While some mistakes may be harmless at times, others can be detrimental in certain instances (3).

When admitted to medical school, students have already performed at their highest level for years (4). In addition to being ambitious and dedicated, they have a thirst for knowledge and spend considerable time studying. They attend lectures, read textbooks, learn the structures and processes in the body, practice examination techniques, recognize symptoms, and learn appropriate treatments. Additionally, they receive clinical training on patients and learn about communication. The fear of making mistakes due to lack of knowledge is present for some during their studies and in their career as a physician (5, 6).

Medical schools guide students along the path from being students to becoming professionals. The professional role is one of the seven roles that new physicians should undertake (7). Understanding, teaching, and assessing what the professional role entails are complex and challenging tasks (8, 9). Epstein et al. defined competence in medicine as "*the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individuals and communities being served*." (10). "*Competence is not an achievement but rather a habit of lifelong learning*." (11, 12). To be professional, physicians must be able to identify their learning needs and respond to them (11). This ability should be instilled during medical school training, where assessment plays an integral role in facilitating this process (11).

Do we train students to do this? All Norwegian medical students take the same final written exam; therefore, the quality of their medical knowledge can be easily compared. However, in 70% of in-hospital adverse events, shortcomings in non-technical skills (NTS) performance have been identified as a contributing factor (13). NTS are not merely medical knowledge but cognitive, social, and personal resource skills that complement the medical knowledge (13). Immediately after graduation, new physicians are expected to handle the roles of independent caregivers and team leaders in emergencies (7, 14). Studies have shown that new physicians often find their new role challenging, as medical school curricula lack comprehensive training for this role (15, 16).

In 2017, a new regional medical education program—*The Finnmark model*—was introduced in the northernmost county of Norway (17). The program focuses on training students to master their roles. It includes extensive use of multi-professional simulation training to equip students with the necessary NTS and prepare them for their professional roles (Figure 1). This thesis is a follow-up study of *the Finnmark model of medical student training* and aims to evaluate the effects of NTS training.

Figure 1. Team training in Finnmark. Photo: UiT - The Arctic University of Tromsø

2 Background

2.1 From student to professional

This thesis focuses on the process of becoming a professional physician rather than a medical student. This process commences on the first day of medical school and culminates when the physician retires or ends clinical practice. Defining the professional role is challenging (9). It has been articulated as follows: “committed to the health and well-being of individual patients and society through ethical practice, high personal standards of behavior, accountability to the profession and society, profession-led regulation, and maintenance of personal health.” (18).

Despite this definition, individuals aspiring to be professional physicians must embrace a lifelong commitment to learning. New examinations and treatments continually emerge, necessitating ongoing education. Competence is inherently context-dependent, even when operating within the same professional setting. The environment is constantly changing. As society is constantly changing, you, your colleagues, and your patients are affected. The ability to adapt to these changes and improve overall performance is part of being professional (11, 19).

Novice medical students are starting out learning rule-based formulas, and further on in the study they are practicing the use of these formulas in different ways in different situations. In residency, they must make decisions based on the specific situation of the patient and a deeper understanding of the underlying principles. Throughout their professional careers, physicians must be able to make rapid decisions in situations with a great degree of uncertainty and effectively communicate their decisions in a way that is adapted to collaborating with healthcare personnel, patients, and their next of kin (11).

2.2 Non-technical skills (NTS)

2.2.1 Definition

The term “non-technical skills” has been debated (20). One reason for the debate is that the term describes something based on what it does not do (21). Alternative terms, such as “human factors skills” and “behavioral skills,” have been suggested. Despite this ongoing debate, we chose to use this term because it is the most common employed phrase in the research field (21).

Flin et al. define NTS as “the cognitive, social and personal resource skills that complement technical skills, and contribute to safe and efficient task performance” (22). These include skills in communication, teamwork, situational awareness, and decision-making (23, 24). In this context, technical skills refer to the profession-specific competence that each individual or occupational group possesses to perform their respective tasks and functions (23, 24). Figure 10 shows examples of NTS (23, 25).

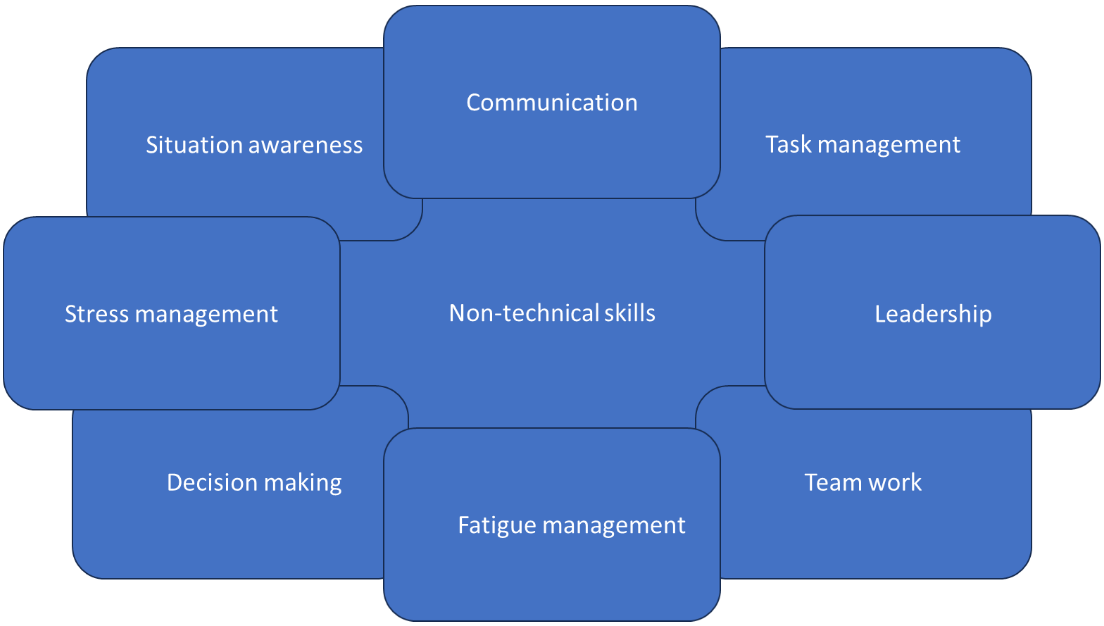


Figure 2 Examples of non-technical skills

2.2.2 Origin

NTS originated in aviation in the early 1970s (26). This began when aviation incidents were first researched. Researchers searching for technical faults in aircraft found that NTS, as communication, could be primarily responsible for accidents (26). As technology has improved, the importance of human error has become more apparent (26). "Cockpit Resource Management" was implemented to reduce the risk of human error(26). This is a set of training procedures to improve safety, focusing on NTS such as communication, leadership, and decision-making (26). This term was later generalized to Crew Resource Management (CRM) (26). Studies have shown that CRM training increases aviation safety (26-28). Other high-risk industries have followed aviation's lead and implemented CRM training, such as ship handling, firefighting, air traffic control, and healthcare (24, 29-32).

2.2.3 NTS in medical education

At the beginning of the 21st century, a paradigm shift occurred in medical education (33). The change from a structure- and process-based system to competency-based medical education (CBME) was initiated by the American Board of Medical Specialties (ABMS) and Accreditation Council for Graduate Medical Education (ACGME) (33). Frank et al. defined CBME as "*An approach to preparing physicians for practice that is fundamentally oriented to graduate outcome abilities and organized around competencies derived from an analysis of societal and patient needs. It de-emphasizes time-based training and promises greater accountability, flexibility, and learner-centeredness"* (34). In other words, there was a shift in the curriculum from time-based to competency-based, and the assessment paradigm changed from knowledge testing to the assessment of competencies (35). Hamza et al. described it as a shift in "*focus away from primarily medical knowledge to envision physician training that* *also emphasizes patient-centered care, communication, professionalism, critical thinking, teamwork, advocacy, and appropriate use of limited resources."* (36).

The importance of NTS competencies has been increasingly acknowledged during the shift to CBME (28). The 2005 updated CanMEDS competency framework identified seven roles that new physicians were expected to master. The seven roles were as follows: Medical Expert, Communicator, Collaborator, Manager, Health Advocate, Scholar, and Professional (7). Medical students are expected to master the roles of independent caregivers and team leaders immediately after graduation (12). This requires high-level performance in technical and non-technical skills. In 2009, the British Parliament mandated training in NTS, even in pregraduate education (37). Studies showed that new physicians often exhibit inadequate NTS performance (15, 38). The shift to CBME has led to new curricula that focus on NTS performance (34). This ongoing reform is described in a Lancet report (39) as follows: "*competency-based education with cross-professional collaborative learning and teamwork, as well as more integration between theoretical and clinical learning in the study courses, technology-supported learning and systems understanding and management. The authors believe this will be a good basis for meeting the needs in the global health service of the future. They also emphasize social accountability both locally and globally" (40).*

2.2.4 Training NTS

Fortunately, it is possible to train and improve NTS performance (41-44). Various educational methods have been reported to have positive outcomes in improving NTS skills and patient safety (45). These include different variants of role playing, interactive lectures, problem-based learning variants, e-learning, computer-based and practical games (46), discussions, didactic presentations, teamwork exercises, group work with medical error cases, video-taped simulations, and simulation-based training (45).

Simulation-based team training is one of the most common methods of NTS training (45). The time and quality of debriefing and feedback are criticaal aspects of the learning process (47). Using an inter-professional team for training is beneficial (24). The multidisciplinary approach increases students' learning outcomes and prepares them for their careers in the healthcare system. Multidisciplinary teams have been established as best practice in many cases across healthcare disciplines (48). Students also described team training with other health profession student groups as valuable. (49)

Multi-professional team training and simulation are recognized as important activities for improving patient safety. A governmental Norwegian Official Report from 2015 called for an increased use of multi-professional training and simulation in healthcare education to improve patient safety (50). In 2015, UiT - The Arctic University of Norway (UiT)'s healthcare education program started a pilot inter-professional simulation training called *InterSim* (51). Medical, nursing, and radiography students in their last year participated in this pilot study. Students evaluated the pilot, and the university received positive feedback. The InterSim program was implemented as a mandatory full-day simulation for students in their final year of medicine, nursing, radiography, and biochemistry studies (52). The students collaborated in inter-professional teams and encountered realistic acute medical scenarios. These scenarios were facilitated by experienced nurses and physicians. Usually, an experienced nurse or physician also played the role of a patient; nonetheless, during the COVID-19 pandemic, it was necessary to use a manikin.

2.2.5 Measuring NTS

The importance of the NTS underlines the necessity of being able to measure it (53). Measuring the NTS is vital to ensure that medical students receive high-quality training, enhancing their NTS performance (54). Students require specific feedback and an evaluation of their strengths and weaknesses in NTS performance in order to more fully develop those skills (54).

CBME, similar to NTS training, requires high-quality, continuous, and frequent assessments (54). There are two main types of assessment: formative assessment *for* learning, and summative assessment *of* learning (45). Carraccio et al. (42) argued that "*a competency-based education program emphasize formative over summative assessment*" (34). The need for summative assessment is indisputable, as there is an obligation to ensure that physicians have sufficient knowledge and competency for unsupervised practice (34). The focus on formative assessment is supported by educational theory (54) and the work on "deliberate practice" (34, 55). Ericsson et al. described the need for effective, high-quality practice to improve skills (55). Clearly, skill goals and regular, immediate feedback on performance are necessary for improvement (55). Professional development depends on effective feedback (56). Feedback is considered one of the most important tools for helping learners progress (56). However, if feedback originates from an inaccurate assessment, it is ineffective (34). Holmboe et al. argued that *"an effective CBME system must continuously link robust assessment with equally robust feedback on a continuous basis"*. Some even consider the "objectification" of assessment unnecessary, as the assessment could be elaborated clearly in words instead of numbers (57). This would also be helpful for students to develop their skills further.

The development of NTS performance includes reflection and evaluation (22). It is important to have a framework of necessary NTS for Norwegian medical students to have a common language for description and discussion. A common language can contribute to increased awareness and recognition of skills (41). Tools are available for several health professionals and practice domains, but not specifically for Norwegian medical students (55-59). To ascertain whether medical students achieve the necessary NTS, an assessment tool is required.

Such a tool can be used during the learning process to provide formative feedback; moreover, it can be used in a summative format to assess the abilities of learners.

2.2.6 Other NTS tools

Several NTS tools are accessible to healthcare professionals (55-59). A systematic review conducted in 2019 revealed 76 unique observer-based NTS assessment tools for health professionals (60). Most assessment tools were designed for multi-professional teams (49%), while 10% were designed for healthcare students (60). Although most tools have been innovatively developed, studies have demonstrated that the development process varies (60). The recommended training for using these tools differs from minimal training (61, 62) to the research groups' own training course (63, 64). The systematic review did not conclude with one recommended tool, but did conclude that a standardized approach to the development of tools would benefit educators and researchers (60).

NTS assessment tools are available for surgeons, anesthesiologists, nurses, and entire teams (55-59, 65). NOTECHS (62), NOTSS (63), and ANTS (64) are among the most well-known. In 2018 the first tool for assessing medical students' non-technical skills were published in Scotland (58). Studies have shown that different health professions and countries have distinct cultures and working conditions that necessitate different NTS (56, 57). Hence, it is essential to develop new tools or adapt existing tools for specific health professions in all countries. Ongoing studies are in progress where more generic tools are being developed for health professionals, but those tools have yet to be presented.

2.2.7 Validation of NTS tools

Studies have shown significant variations in the validation of NTS assessment tools for healthcare professionals (60). A systematic review in 2019 showed that 80% of the tools were validated using inter-rater testing or internal consistency (60). This review found few pragmatic assessments of tool usability (60). Another systematic review evaluated the validity of simulation-based assessment scores in 217 articles using two or more sources of evidence (53). Only 3% of the articles referenced the standard framework, and 24% did not reference any validity framework. In most articles (94%), the relationships with other variables were calculated, most often relating to the training level. Evidence of validity for internal structure (76%) and content (63%) was evaluated in most studies. Evidence of the validity of the response process and consequences was present in less than 10% of studies.

When collecting evidence of validity for NTS assessment tools, interrater reliability is often referred to as the most valuable factor (66). This measures the level of agreement between the raters and is often determined by calculating the intraclass correlation coefficient (ICC) (66). In 2013, the *"Guidelines on the Requirements of a 'Train-the-Trainers' Program to Assess Nontechnical Skills"(67)* suggested the following ICC levels without reaching an overall consensus:

- ***To provide performance feedback within clinical practice:*** *Minimum ICCs of 0.61–0.70*
- ***To carry out assessments for research purposes:*** *Minimum ICCs of 0.71–0.80*
- ***To provide high-stakes assessments:*** *Minimum IICs of 0.71–0.80*

Some studies suggest that summative high-stakes assessments should always be conducted by at least two assessors (54, 65). Higham et al. examined the validity and usability of the ANTS, NOTECHS, and OSCAR using three raters (65).

The 2013 guidelines (67) also considered a half-day training program to be practical but found a two-day training program with post-training support and a refresher course to be more desirable. However, experts have not reached a consensus on the necessity of calibration exercises (67). This was attributed to the lack of time for such a calibration. There is a considerable gap between the desired competencies of NTS tool users and the time available to train busy physicians as near-peer educators. Despite the increasing use of simulations, most medical education programs are conducted in clinical settings. Learning and assessment will mainly occur in clinical workspaces in the future (54). Holmboe et al. argues that "*Traditional approaches to measurement, based in the psychometric imperative, have been leery of work-based assessment, given the biases inherent in the clinical setting and the challenges of 'adjusting' for contextual factors that make it difficult to determine the 'true' score, or rating, of competence. The implications of these considerations for CBME is that this approach to medical education must account for and incorporate contextual factors arising from the clinical setting into assessment processes."* (54).

Variations in the development, validation, training, and use of NTS tools impede their usability. Standardization of NTS assessment tools and training may be helpful for healthcare students and educators (38). The assessment of these important skills should be accessible, standardized, simplified, and supported by adequate skills (38). It is recommended that all medical education programs in a country use a core set of assessment tools (54).

2.3 Patient safety

Norway's public healthcare system generally maintains high levels of patient safety (68), which is attributed to several factors (69). Norway's financial situation facilitates high-quality health services (69). Moreover, the country provides high-quality healthcare education (69). Over the last 15–20 years, there has been an increasing focus on patient safety and quality (70). Various projects have been initiated at the national, regional, and local levels (70). These initiatives include patient safety programs, national guidelines, safety standards, standardized patient care pathways, the project “One Citizen — One Record” (71), systematic prevention of healthcare-associated infections, and the reform “Live your whole life” (72)—a quality reform for the elderly population (70). Regular national patient safety campaigns and learning programs are available for health professionals to raise awareness (70). The quality and safety of the healthcare system are monitored using national quality indicators (70). National and local standardized guidelines and procedures have been developed and implemented to reduce the risk of error (70). Furthermore, there are different local and national reporting systems for adverse events where healthcare professionals or patients can report errors (70). The idea is that errors should be reported to identify risks and areas of improvement and learn from them. In 2017, regulations on *Management and quality improvement in the health and care service* were passed (73). In 2019, the *National Action Plan for Patient Safety and Quality Improvement* was launched by the Norwegian Directorate of Health (70). The goal of this plan was to *“Reduce the proportion of patient stays with at least one patient injury, all degrees of severity"*, by 25% from 2017 to the end of 2023 (70). This corresponds to a reduction from 13.7% (2017) to 10.3% (70, 74). In 2022, despite all initiatives, it was reduced to only 12.7% (68).

The Norwegian System of Patient Injury Compensation (NPE) handles claims from patients who believe they have suffered damage due to errors in the healthcare system (75). It is a government agency under the Norwegian Ministry of Health and Care Services (75). Every year, NPE pays approximately one billion NOK in compensation for patients owing to incorrect treatment (75). Reports have shown a slight downward trend from 2012 to 2021 ($p = 0.102$) (74).

The reason for the lack of effect of these error reducing initiates can be found in international studies. Patient safety is a topic of significant interest worldwide. The World Health Organization launched the World Alliance of Patient Safety in 2004 (76). Subsequently, there has been an increasing focus, and several programs and activities have been initiated at various levels. However, the effects of these patient safety programs and interventions remain unclear (77).

Studies show that counting errors and the bureaucratization of safety work have counter-effects (78). Each program may work as part of the Norwegian *National strategy against antibiotic resistance 2015-2020*, which achieved the goal of reducing the use of antibiotics in Norway by 30% (79). Nevertheless, the bureaucratization of guidelines, programs, and the measurement of quality indicators to prevent human errors will have secondary effects (80). Studies have shown that if primary care physicians follow all relevant guidelines, they would have to work 27 hours a day (81). Therefore, if physicians follow all the rules, they will not be able to provide patient care, and adverse events will occur.

Human factors and ergonomics (HFE) have been shown to improve patient safety. It is recommended as a key systems engineering tool for improving the quality of care and patient safety (82). There is a need to increase the use and implementation of HFE tools and methods in healthcare systems to improve patient safety (83). Human factors include an understanding of human capabilities, behaviors, and limitations in optimizing systems for human use (84). This approach is particularly crucial in high-risk industries, such as aviation and healthcare (83). Human factors in these industries include NTS, such as interpersonal and cognitive abilities (85). Implementation of HFE tools in the healthcare system includes improving NTS.

In OECD Health Working Paper No. 96, titled "*The economics of patient safety: Strengthening a value-based approach to reducing patient harm at national level*" from 2017, it was found that investing in fundamental long-term programs in professional education and training is necessary for improving patient safety (86). The authors describe professional education as a skill that goes beyond technical and clinical skills (86). Important sociotechnical competencies include communication, teamwork, and self-reflection, and socialized healthcare professionals can establish inter-professional relations—these are consistent with what is typically described as NTS.

2.4 Decentralized education

Traditionally, medical schools have been centered around large university hospitals in cities (87). The distribution of health professionals has also been skewed toward urban areas. The transformation of medical education to CBME has increased attention to the important development of the necessary skills and competencies required by patients and the community (29-31). A paradigm shift is underway in several countries where more suitable learning environments have been identified (87)—training platforms that help health personnel become socially accountable (88) and "fit for practice" (89). This paradigm shift has led to an increased focus on the need for medical education in primary healthcare, especially in rural and remote areas (87). The aim is to produce healthcare professionals who are appropriately trained to work in rural areas (90). Studies have shown that health professionals trained in decentralized training platforms (DTPs) are more likely to stay in rural areas (91-93).

Studies have found that DTPs enhance student learning in contextual environments (87). During the years of CBME, the importance of transformative learning has increased. Mezirow introduced transformative learning as a theoretical construct in 1991 (94): "*learning that transforms problematic frames of reference—sets of fixed assumptions and expectations (habits of mind, meaning, perspectives, mindsets)—to make them more inclusive, discriminating, open, reflective, and emotionally able to change*" (95). Transformative learning and social accountability are inseparable. Therefore, DTPs are the preferred learning environment for transformative learning (87). Community engagement also contributes to successful learning (87). The need for "*education in the 21st century that fosters adaptation* *of professional competencies that are specific to the rural context*" was highlighted in the 2010 Lancet Commission report on health professions education (39). The report recommended that DTPs be located close to rural communities. Simulation-based NTS training also facilitated transformative learning (96).

2.4.1 The Finnmark model

Norway is a geographically elongated country facing demographic challenges. Centralization and a demographic shift to urban centers poses difficulties in recruiting professionals to rural areas (29). Healthcare professionals and physicians are no exceptions (28, 29). Several projects involving decentralized education by physicians have been initiated in Norway (30, 31).

This thesis is one of two follow-up studies on decentralized medical education in Norway and *The Finnmark Model for Medical Student Training*. This is a decentralized model in which a subset of medical students from UiT undertake the last two years of medical school in Finnmark County. The first four years of the study are at the main campus in Tromsø, where the University Hospital of the region is situated. Finnmark County is the northernmost county in Norway. It is a rural area covering 48,600 km2 and has a population of approximately 75,000 people. The county has two hospitals: one in Hammerfest with 100 beds and one in Kirkenes with 50 beds. There are several decentralized medical and mental health services in the county.

The Finnmark model was designed to educate physicians with the necessary medical skills to serve in rural areas. The students were expected to attain all aspects of clinical skills in two years in Finnmark. In addition to the main campus in Tromsø, there were decentralized students during the fifth and sixth years in Bodø. Students at all three sites had the same learning objectives. Learning activities and settings for achieving the required skills differed.

In Finnmark, students engage in learning activities in cultural diversity and understanding, especially toward the indigenous Sami population. The Finnmark Model also focuses on emergency medicine and the interaction between primary and specialist healthcare services. Finally, the Finnmark model continuously uses simulations with a focus on NTS performance (Figure 3).

All the students completed the same final exam, which comprised one written examination and four oral examinations. All five examinations were were graded on a pass/fail basis. If one fails, they must take all five exams again after six months. Since 2021, written examinations have become a common national exam for all medical faculties in Norway. Hence, technical skills at different training sites can be compared in the final exam.

Figure 3 Team training at The Finnmark Model UiT Photo: UiT

2.5 User perspective

Norway requires user participation in health research. User participation in the research is part of the commissioner's Document from the Ministry of Health and Care Services to the Regional Health Authority. National guidelines for user participation were established in 2015 (97). In 2018, a national guide for user participation in health research by specialist health services was launched (98). This tool aims to clarify the guidelines and make them more accessible.

The aim of user participation in health research is to enhance the relevance and quality of the research. If research results prove beneficial to patients and their relatives, they are more likely to be implemented quickly and are more significant (98).

The guidelines state that user participants should be involved in planning and conducting research projects. User participants primarily comprised patients and their relatives (Figure 4). Ideally, they should be represented by patients or user organizations. In some cases, healthcare professionals or the general population may also be considered users of health research (98).

Figure 4 User committee at the Finnmark Hospital Trust. Photo: Finnmark Hospital Trust

3 Aims

This project was a follow-up to the *Finnmark Model of Medical Student Training (99)*. We wanted to determine whether the decentralized students in Finnmark (Figure 5) became good physicians in terms of achieving NTS, in addition to factual medical competencies. Extensive training in NTS improves NTS performance. To achieve this, a tool is required to assess Norwegian medical students' NTS. This tool could be used to compare NTS performance between students at the same university with similar learning objects and curricula but different learning activities on different campuses.

The aims of the thesis were as follows:

- Find the necessary NTS for newly graduated Norwegian physicians
- Develop a usable tool with evidence of validity for the assessment of Norwegian medical students' NTS
- Use the tool to compare NTS levels among medical students at different training sites

Figure 5. Medical students training non-technical skills. Photo: UiT - The Arctic University of Norway

4 Materials and methods

The approach chosen for each study was based on the research question (100). A mixed-methods approach was considered the most relevant to reach the overarching goal of this thesis—creating a tool to assess Norwegian medical students' NTS and compare NTS performance levels at different training sites.

Johnson et al. (101) defined mixed methods as follows: "*Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e. g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration.*" In other words, this study utilized a combination of qualitative and quantitative research methods, enhancing the scope and robustness of the investigation (102).

4.1 Paper 1: The development of a tool to assess Norwegian medical students' non-technical skills

The first study was a qualitative and exploratory study aimed at identifying the necessary NTS for Norwegian medical students and creating an assessment tool for these skills (103). Qualitative methods are suitable for exploring the opinions of informants. When creating new tools, it is necessary to explore these opinions. In this study, we conducted focus group interviews. We analyzed the interviews using a modified open-card sort analysis and created a prototype (104). A literature search was conducted after the interviews and analyses to compare the prototype with existing tools and literature.

Focus group interviews were the preferred method for examining the participants' experiences, opinions, desires, and concerns. The focus group interviews utilized group dynamics to help participants identify and clarify their views (67). This approach is an effective way to explore the participants' perceptions at a profound level. All interviews were conducted by the same researcher, who also played the role of a moderator. This led to uniform moderation and conducting of the interviews. It was also easier for the moderator to determine when data saturation was reached. After five interviews, saturation was deemed to have been reached, and two additional interviews were conducted to confirm this. All interviews were audio-recorded and transcribed.

The transcribed interviews were subsequently transferred to NVivo (alfasoft.com) and analyzed using a modified open card sort analysis (104). This approach was chosen because we had created a new tool and did not have preliminary categories for sorting the units of meaning. Two researchers individually performed the open card sort analysis and discussed the results until a consensus was reached.

After the interviews were completed, a literature search was conducted to ensure that comprehensive scientific works on other types of NTS tools were considered during the development of our tool.

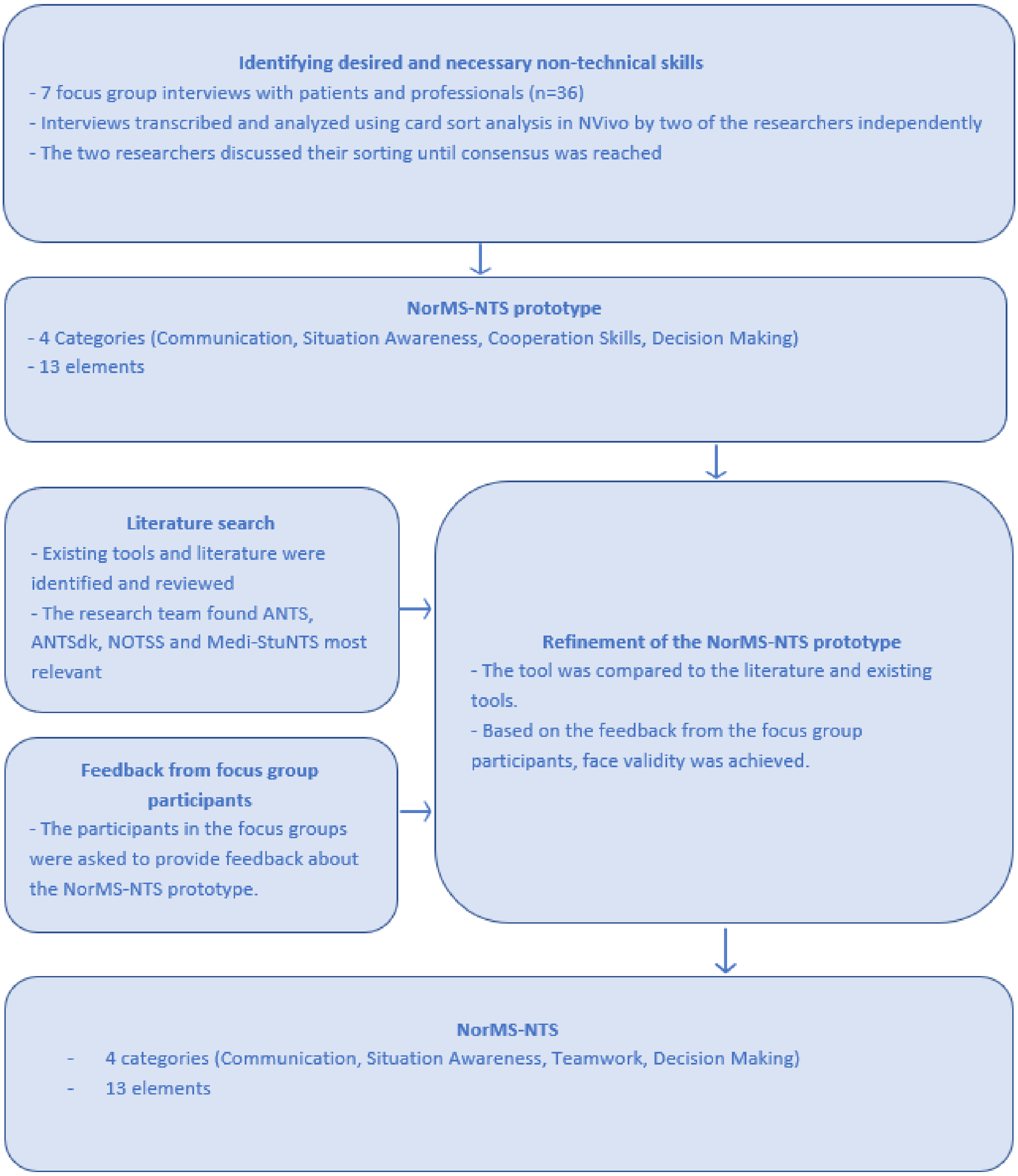


Figure 6 Process of developing NorMS-NTS. Adapted from: Prydz K., Dieckmann P., Musson D. & Wisborg T. (2022). The development of a tool to assess medical students' non-technical skills - the Norwegian Medical Students' Non-Technical Skills (NorMS-NTS). Medical Teacher.

4.2 Paper 2: Collecting evidence of validity for an assessment tool for Norwegian medical students' non-technical skills (NorMS-NTS): usability and reliability when used by novice raters

The second study was a mixed-methods study designed to collect evidence supporting the validity of the NorMS-NTS tool. The validation of assessment tools is complex and context-dependent (105). It is not the tool itself that is validated but the use of the tool in a specific setting (105). We aimed to validate a tool used by novice raters for formative assessment.

Three raters rated 20 videos of medical students acting as team leaders in multi-professional team training, utilizing the NorMS-NTS after a brief introduction. Subsequently, we conducted quantitative research by calculating the interrater reliability, internal consistency, and observability.

The qualitative aspect of this study involved a questionnaire distributed to the raters after they completed rating the videos. Participants were asked to provide feedback on the usability of the tool.

4.2.1 Messick framework

There are different frameworks to choose from when collecting evidence of the validity of assessment tools. We opted for the conceptual and theoretical Messick framework, which has been the standard in the field since 1999 (106). This framework seeks to "*gather evidence showing a connection between assessments and specific constructs*" (107). This approach provides a structure for gathering evidence of validity for the NorMS-NTS tool.

The Messick framework comprises five sources of evidence (106):

Content: This is described as "*The relationship between the content of a test and the construct it is intended to measure*" *(106).* This evidence can be collected through item sampling, development, and scoring. It was assessed during the tool's development, as described in Paper 1 (103).

Internal structure: This is described as "*The relationship among data items within the assessment and how these relate to the overarching construct*" *(105).* This type of evidence can be collected by calculating internal consistency reliability, interrater reliability, factor analysis, and test item statistics (105).

Interrater reliability was analyzed in Paper 2 (108). As all participants were rated by the same raters, we calculated intra class correlation (ICC) (3,1) and Kendall's W.

We also conducted an *internal consistency analysis* to explore the correlations between the elements, categories, and overall global scores . Spearman's nonparametric correlation and Cronbach's alpha (CA) were calculated.

We also analyzed the *observability* of each element, category, and overall global score. This is considered acceptable if observability >50%.(109)

Relationship with other variables: This is described as "*The degree to which these relationships are consistent with the construct underlying the proposed test score interpretations*" *(106).* This type of evidence can be calculated by correlation with tests measuring similar constructs, correlation with tests measuring different constructs, or expert-novice comparison (105).

This part of collecting evidence of validity is planned for further studies.

Response process: This is described as "*The fit between the construct and the detailed nature* *of performance . . . actually engaged in*" *(106).* This type of evidence can be collected by analyzing examinees or raters' thoughts or actions during assessment, assessment security, quality control, or rater training (105).

We sent all the raters a questionnaire after completing the ratings. The questionnaire comprised questions regarding the usability of the tools.

<u>Consequence:</u> This is described as "*The impact, beneficial or harmful and intended or unintended, of assessment*" *(105).* This evidence can be collected by examining the impact on examinee performance, board scores, graduation rates, clinical performance, other examinee effects, or the definition of a pass/fail standard (105).

Part of this was conducted in Study 2 (108), as we evaluated the possibility of minimal rater training. Further studies are planned to collect evidence for the validity of these consequences. Table 1 presents the five sources of evidence of validity in the Messick framework and the procedure used in Paper 2 (105, 108).

Table 1. The five sources of evidence of validity and the procedure used in study 2.

Source of evidence	**Definition**	**Procedure**
Content	"the relationship between a test's content and the construct it is intended to measure."(106)	*Assessed as a part of development.*
Internal structure	"The relationship among data items within the assessment and how these relate to the overarching construct."(105)	Interrater reliability Internal consistency Observability
Relationships with other variables	"The degree to which these relationships are consistent with the construct underlying the proposed test score interpretations."(106)	*Planned in further validations*
Response process	"The fit between the construct and the detailed nature of performance . . . actually engaged in."(106)	Raters respond in questionnaire
Consequences	"The impact, beneficial or harmful and intended or unintended, of assessment."(105)	Evaluation of the possibility of minimal rater training

4.3 Paper 3: Non-technical skills of Norwegian medical students at different training sites: A comparative, observational cohort study.

The third study was a quantitative observational cohort study. The purpose of this study was to utilize the NorMS-NTS in a training environment, and to investigate possible performance differences across three different training sites.

We recorded eight videos from each training site for a total of 24 videos. Eight raters used the NorMS-NTS tool to rate the team leaders' (medical students) NTS performance in each video.

Subsequently, we compared the average NTS performance scores among student groups at each training site. The null-hypothesis was no significant difference between training sites studied, and a Tukey's test was conducted to calculate any significant differences (110, 111).

4.4 Video recordings

Raters were blinded to the training sites studied in article 3 and were tasked with comparing student performance at different training sites. We found it reasonable and practical to record videos of medical students at the different training sites. Video recordings used to assess performance have been found to be as reliable as in-person assessments (112). All campuses underwent InterSim training in the sixth year of the study period. This is a simulation-based multi-professional team-training program with the same written scenarios on all campuses. The training sessions were recorded. In the period 2018–2023, we recorded students in their sixth year of medical school in Bodø, Tromsø, and Finnmark and students in their fifth year in Finnmark. Owing to the COVID-19 pandemic, students were not allowed to use InterSim in 2020 and 2021. More than 100 videos of over 500 students were recorded. All the videos were checked for sound and image quality; many were discarded owing to their low quality.

From the videos with sufficient sound and picture, 20 videos were randomly selected for use in Study 2 and 24 were selected for use in Study 3.

4.5 User perspective

This project had a representative from the user committee of the Finnmark Hospital Trust, which was involved in the planning of the entire project. The user committee was also part of the first study, serving as participants in two of the focus groups. In the second study, the tool was tested by the intended users—busy physicians with limited time for training in using the tool.

4.6 Data storage, approvals, and ethical considerations

In this project, we video-recorded the data from more than 500 healthcare students. The multi-professional team training we recorded was compulsory for most students. All students received written information about the purpose and objectives of the study prior to the training and were given oral information on the day of training. They were informed that their participation in the project and recording were voluntary. If a student opted out, the team was not recorded. There were no consequences for the students. Students were informed that the videos would be safely stored and rated by raters who were blinded to their names and places of study. Two students did not want to participate, and their team training was not recorded. The focus group interviews were recorded and stored safely until they were transcribed. After transcription, the recordings were deleted.

Written informed consent was obtained from all the participants after providing oral and written information regarding the purpose and objectives of the study.

All videos were stored and rated on the Tjeneste for Sensitive data (TSD) facilities. The University of Oslo (UiO) owns the TSD, which is operated and developed by the TSD service group at UiO IT Department (USIT) (tsd-drift@usit.uio.no). TSD provides a secure platform for Norwegian public research institutions. The platform provides the ability to collect, store, and process sensitive data.

The research protocol for this thesis was submitted with an application to the Regional Committee of North Norway for Medical and Health Research (Ref: 2016/1539/REK nord). The need for a formal review of the study was waived, as Norwegian law exempts educational studies from medical ethical approval if they do not involve patients.

In 2017, the Data Protection Official for Research for Finnmarkssykehuset was the Norwegian Centre for Research Data (NSD). They approved the interview guide for the focus group interviews, consent form, and data storage (NSD Ref: 57474/2017). The purpose of NSD is “to ensure legal access to necessary personal data for research” (Norwegian Centre for Research Data).

5 Results

5.1 Paper 1

We established the NTS necessary for a newly graduated physician in Norway. Subsequently, we created a tool to assess Norwegian medical students' NTS, the NorMS-NTS (103). The tool is comprised of four categories and 13 elements (Figure 7) of observable behaviours. We also created a handbook to use with this tool (see Appendix).

HELSE NORD | DAVVIDEARVVAŠVUOHTA VARRESVUOHTA NUORTTA HEALSOE NOERHTE

NorMS-NTS **Student nr: __**

Kategori**	Kategori vurdering*	Elementer**	Element vurdering*	Tilbakemelding
Kommunikasjon		Kommunikasjon med team		
		Etablere felles forståelse		
		Kommunikasjon med pasient		
Situasjonsbevissthet		Sette seg inn i situasjonen		
		Forstå ulike roller i teamet		
		Oppmerksomhet		
Samarbeidsevne		Ydmykhet		
		Fleksibilitet		
		Bruke teamets ressurser		
Beslutningstaking		Gjøre gode valg		
		Håndtere usikkerhet		
		Lederskap		
		Prioritering		

Generelle kommentarer:

*N, Ikke observert. 1, Langt under gjennomsnittet; 2, under gjennomsnittet; 3, akseptabel; 4, over gjennomsnittet; 5, langt over gjennomsnittet..
** Innad i teamet om ikke annet er spesifisert.

Samlet vurdering (Sett ring rundt)**:**
Svært dårlig 1 – 2 – 3 – 4 – 5 – 6 – 7 Utmerket

Figure 7 The NorMS-NTS tool

Communication

The first category in the NorMS-NTS is *Communication*, which was important for all focus groups (103). They emphasize the importance of communication as a two-way event that characterizes a good team. It is not merely the team leader's responsibility but also how the recipient perceives what is said.

Others have described communication as a core NTS. (113) Communication as an NTS in healthcare has been defined as follows*: "a means to provide knowledge, institute relationships, establish predictable behaviour patterns, and as a vital component for leadership and team coordination"* (24)

The three elements of *Communication* are *Team communication, Establish mutual understanding,* and *Patient communication.*

Situation awareness

The second category is *Situation awareness.* The participants described this as a requirement for the physician to be present and aware (103). Situational awareness has been defined as *"the perception of elements in the environment, the comprehension of their meaning in terms of task goals, and the projection of their status in the near future"* (24, 114). This is a three-level process: perception (gathering relevant information about the situation), comprehension (the ability to form a differential diagnosis), and projection (the ability to make a prognosis) (24).

Through the development process, we identified three elements in this category. *Situational assessment* is the ability to gather, analyze, synthesize, and communicate data. *Understanding of team members' roles* is necessary to understand the environment and gather information. *Attentiveness* is important for gathering information.

Teamwork

The third category was teamwork. Here, the informants described the importance of being aware of other professions' knowledge and competencies and delegating relevant tasks to create an efficient team (103). They also noted respect as a keyword and found that newly graduated physicians must respect the competence and work experience of other health professionals.

Salas et al. defined a team as *"a distinguishable set of two or more individuals who interact dynamically, adaptively, and interdependently; who share common goals or purposes; and who have specific roles or functions to perform."* (115) Teamwork depends on all the contributions of team members. In our tool, we scored the medical students' skills necessary to interact dynamically, adaptively, and interdependently — *professional modesty,* awareness of other team members' knowledge and experience, and the student's limitations. Things change rapidly, and the ability to adapt quickly and exhibit *flexibility* is important. Competency refers to the *efficient use of team members.* Good teamwork necessitates that all members contribute according to their strengths.

Decision making

The fourth category is Decision making. The focus groups were all concerned with uncertainty management (103). Both the physician's own lack of experience, and the uncertainty of the patient's situation. While all tests inherently involve some degree of uncertainty, it is not always possible to wait for test results before making a decision. One participant described this as a distinctive feature of working in general practice, emphasizing that many decisions are based on uncertainty. Physicians also often face their first encounters with handling uncertainty in general practice. It is necessary to develop these skills to make decisions under time pressure and on an uncertain basis. Few existing NTS assessment tools incorporate uncertainty as part of their tool. However, the only existing tool for assessing medical students' NTS, the Medi-StuNTS (58), addresses uncertainty.

Others have described decision-making as follows: "Assessing the situation and making a decision. Communicating plan(s) and implementing decisions" (56). The inputs from the focus groups constituted the following four elements: *"Uncertainty management," "Decision analysis," "Leadership," and "Prioritization*." Medical students must evolve their *Uncertainty management*, as well as the ability to conduct *Decision analysis* to make informed choices under uncertainty. Communicating and implementing decisions through *Leadership* are also important. During decision-making, *prioritization* is also necessary regarding which examinations and treatments are performed and when.

In the other tools, "Leadership" was one category, but the leadership the informants described was the ability to make the right decisions at the right time. Hence, it has become an element in decision-making instead of an individual category.

5.2 Paper 2

In study 2, we collected evidence for the validity of the NorMS-NTS tool (108). We found sufficient interrater reliability, internal consistency, and observability for formative assessments. The raters found the tool to be usable. These results are widely described in Paper 2. In summary, the results are as follows.

Internal structure:

Interrater reliability: We found fair ICC agreement for all raters for the sum score of the overall global score: ICC (3,1) = 0,53 (66). This was supported by Kendall's W = 0,73.

Internal consistency analysis: Most Spearman's correlations were above 0.80, and Cronbach's alpha was mostly above 0.90. A value of 1.0, represents high internal consistency, so both were considered sufficient.

Observability: The observability varied from 95% to 100% and was deemed acceptable.

Response process:

The raters found the tool easy to use.

The results of Study 2, according to the Messick Framework, are summarized in Table 2.

Table 2. A summary of the results in Study 2, According to Messick framework

Source of evidence	Definiton	Procedure	Results
Content	"relationship between a test's content and the construct it is intended to measure."	*Validated during development process*	Not evaluated in this study.
Internal structure	"Relationship among data items within the assessment and how these relate to the overarching construct"	Inter-rater reliability	*Sufficient*
		Internal consistency reliability	*Sufficient*
		Observability	*Sufficient*
Relationships with other variables	"Degree to which these relationships are consistent with the construct underlying the proposed test score interpretations"	*Planned in further validations*	Not evaluated in this study.
Response process	"The fit between the construct and the detailed nature of performance . . . actually engaged in"	Raters response in questionnaire	*Sufficient*
Consequences	"The impact, beneficial or harmful and intended or unintended, of assessment"	Evaluate the possibility of minimal rater training	*Sufficient results of minimal rater training. Necessary with further studies of consequenses.*

5.3 Paper 3

In Paper 3, we compared medical students' NTS performance at three different training sites—Finnmark, Bodø, and Tromsø—using the NorMS-NTS tool. We hypothesized that there would be no significant difference. We found no significant difference between Bodø (mean 3.5) and Tromsø (mean 3.8), but Finnmark had a significantly greater level of overall NTS performance (mean 4.5).

Similar results were obtained at the category level. Finnmark participants had significantly higher levels of NTS performance than those at Bodø and Tromsø. With the exception of only one category, there were no significant differences between participants at Bodø and Tromsø. In the category "Decision-making," Bodø participants scored significantly lower than those from Tromsø.

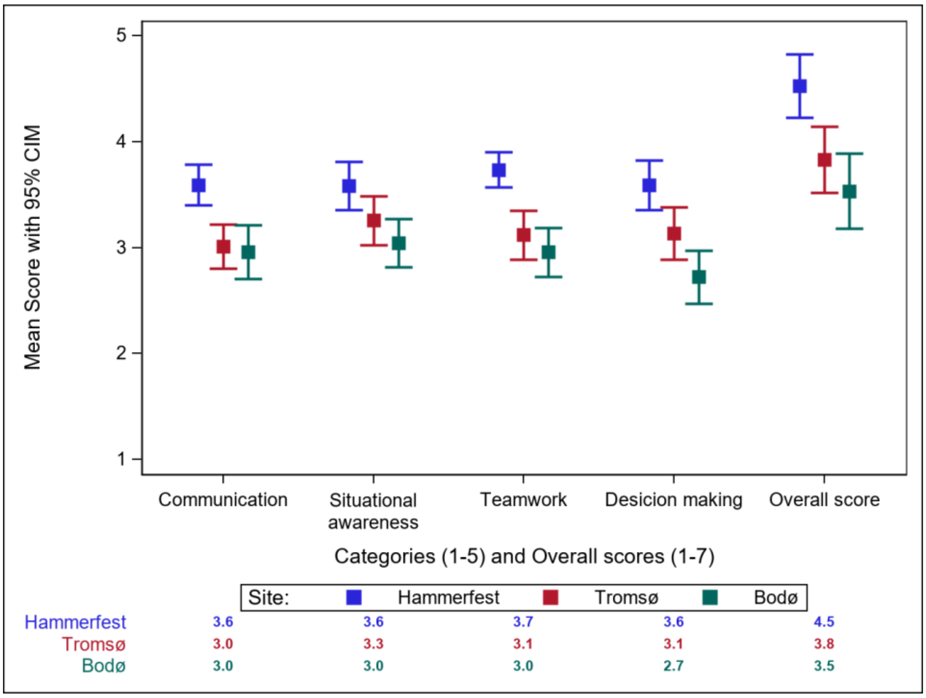


Figure 8. Overall and Category scores for NTS performance of different training sites. CIM = Confidence Interval of the Mean

6 Discussion

In this thesis, we defined the necessary NTS for Norwegian medical students (103) and created an easy-to-use tool to assess these skills in near-peer training. We utilized this tool to examine the student performance at decentralized training sites by comparing NTS scores within student groups with the same learning goals and activities. Our findings reveal that Finnmark students had significantly higher levels of NTS performance than the Bodø and Tromsø students.

6.1.1 Adapt or create

To create a tool for assessing NTS, we needed to identify the essential NTS in this setting. We opted to conduct a qualitative and exploratory study to extend our knowledge of medical students' NTS (116). The research group discussed whether to use an existing NTS tool and customize it or develop a new tool from scratch. Customizing an existing tool would preserve the comprehensive scientific work behind it, and creating a new tool from scratch could prevent overlooking specific NTS for this group; however, it also requires additional work and time. Higham et al. described the development of 76 NTS tools for healthcare professionals (60). They found that the developmental process varies widely (60). There is no standardized approach to the development of NTS tools (60). We adopted a broader approach for creating a new tool. After creating the NorMS-NTS tool, we found it to be similar to existing tools (103). This finding supports the content validity of the NorMS-NTS and existing tools. This finding also facilitates future research on adapting existing tools rather than creating new ones.

6.1.2 Evidence of validity

Higham et al. reported that assessments of the validity of NTS tools varied widely (60). They stated that evidence of the validity and usability of different NTS tools is important. Research on usability has been lacking for several years, and most related studies have focused on tools developed over the last ten years. This article describes the importance of considering usability, especially because of the increasing shortage of healthcare staff. The distinction between formative tools requires minimal training and can be implemented easily. Formative tools can be used for training and rewarding feedback. However, high-stakes assessment tools with robust validity and reliability can only be utilized by experienced raters (57). When the research group selected the evidence of validity that we prioritized collecting first, we chose what we deemed most important. To achieve the highest level of utility, the tool must be widely used. Norwegian medical students are often taught by busy clinicians. We wanted the tool to function in this setting. Thus, we could not create a tool that required a two-day course. A cross-sectional study among facilitators observing behavioral skills in healthcare found that less than 50% of facilitators had formal training in the framework they used for observation (117). Hence, we chose to collect evidence for the validity of the NorMS-NTS tool for formative assessment when used by novice raters.

6.1.3 Messick framework

The validation of an NTS tool is not a simple task (56). Using these tools in different settings may yield divergent results. Score, interpretation, and use affect validity. When collecting evidence of validity, it is necessary to specify the intended use and settings of the tool. The validation of these tools is a continuous process of collecting evidence of their validity in different contexts. Evidence of validity collected in a specific setting may be transferable to another setting but always requires individual consideration.

As the validation process is continuous over time, it is necessary to follow this process to obtain a transparent map of validated and non-validated tools. This ensures that other researchers and educators understand how to use the tool and identify missing pieces for further research. Hence, standardized frameworks are used (60).

The studies in our project can easily be categorized within the correct source of evidence in the Messick framework.

Content evidence was collected from Study 1 (103). We conducted focus group interviews with people working with or treated by newly graduated physicians. After analyzing the interviews and creating a prototype of the tool, we asked the participants in the groups for feedback. Participants confirmed that the tool reflected their opinions and inputs. We also found that the tool was similar to existing tools, thus supporting its content validity.

Internal structure evidence was collected from Studies 2 and 3. We calculated interrater reliability and observability and conducted an internal consistency analysis.

Relationships with other variables are planned in future validations.

Response process evidence was collected in Studies 2 and 3. All raters participating in these studies received a questionnaire after they had rated the videos. The raters provided feedback on the tool and its use.

Consequence evidence was partly collected by evaluating the possibility of minimal rating training in Study 1.

We could have used another framework, but the Messick framework offered an excellent overview of the evidence of validity. We collected evidence from four of the five suggested sources. As validation is a continuous process, further validations are necessary, and the framework provides a good overview of the types of evidence that have already been collected and what should be further investigated.

6.1.4 Observation of NTS

In a 2023 study, Mommers et al. reported that observing NTS is a complex and challenging task, even with high-quality assessment tools (117). They found three major causes of the difficulty in observing NTS (117): "*1. Not everything can be observed*:" This limitation may stem from the script of the scenario or because certain aspects are explicitly expressed, such as mental models; "*2. Not everything is observed*:" There is always a possibility for the observer to miss the expressed NTS; "*3. Interpretation of the observed NTS is difficult:*" The interpretation of observed skills always involves a "subjective process" influenced by the raters' experiences and references. Although tools describe the necessary NTS, they are not clear, objective standards. The performance of NTS may also fluctuate during a scenario, and technical skills and NTS may be intertwined, affecting interpretation. Despite these difficulties, 95% of the informants found the frameworks useful (117).

The wide range of methods for collecting evidence of validity for NTS assessment tools, mostly validated using the average measure of ICC, reflects complexity (65). To reach sufficient levels of ICC despite difficulties with objective observation, it is necessary to compare at the group level instead of at the individual rater level. We consider the possibility of developing an NTS assessment tool with an excellent single-measure ICC and clear objective standards as unattainable goals.

In Paper 2, we collected evidence for the validity of NorMS-NTS using a single-measure ICC (108). We discovered that the tool exhibits sufficient interrater reliability, internal consistency, and observability for formative assessment when used by novice raters (109). Raters found the tool relevant and easy to use for training and teaching (109). However, the scores are far from excellent (66). If we calculated the average ICC, we would have achieved a higher value (118). Nevertheless, the tool demonstrated required features such as inter-rater reliability between individual raters and is easy to use. We consider our findings to be relevant and important features of the tool and that the values are at the expected levels (63, 119).

Nonetheless, further evidence of the tool's validity is necessary. The consequences of using this tool are particularly significant. What impact does the tool have on students? Is it possible to implement the tool? Will the implementation of the tool increase students' NTS performance? These important questions must be addressed in future studies.

After finishing the second paper, we became aware of the Guidelines for Reporting Reliability and Agreement studies (GRRAS) (120). The authors identified a lack of widely accepted guidelines for reporting reliability in the medical field. However, the level of reliability among tool users remains unknown because studies on reliability have failed to report adequate results. Therefore, the interpretation and synthesis of the study results are difficult. The authors of the GRRAS paper designed a guideline for 15 issues that should be addressed when reliability and agreement are reported in the health and medical fields. We analyzed Paper 2 using these guidelines. The analysis results are summarized in Table 3.

We found that we followed the guidelines to a large extent, but we did not explain how the sample size was chosen. Hence, we describe and discuss the sample size under the methodological considerations.

Table 3. Results after analyzing paper 3 using the Guidelines for Reporting Reliability and Agreement studies (GRRAS)

Title and abstract	1. Identify in title or abstract that interrater/intrarater reliability or agreement was investigated.	*Yes. The abstract descibes that interrater reliability was investigated.*
Introduction	2. Name and describe the diagnostic or measurement device of interest explicitly.	*Yes. The NorMS-NTS tool is described explicitly.*
	3. Specify the subject population of interest.	*Yes. The subject population of interest, the Norwegian medical students, are described.*
	4. Specify the rater population of interest (if applicable).	*Yes. The rater populatione of interest, novice raters, are described.*
	5. Describe what is already known about reliability and agreement and provide a rationale for the study (if applicable).	*Yes. Validation as a continous process is described, and that this was the first part of the validation process of the tool.*
Methods	6. Explain how the sample size was chosen. State the determined number of raters, subjects/objects, and replicate observations.	*It is not described how the sample size was chosen. The numbers of raters (3), subjects/obects (20) and replicate observations are described.*
	7. Describe the sampling method.	*The sampling method are described, but not named.*
	8. Describe the measurement/rating process (e.g. time interval between repeated measurements, availability of clinical information, blinding).	*Yes. The rating process and blinding is thorougly described.*
	9. State whether measurements/ratings were conducted independently.	*Yes. It is stated that the ratings were conducted independently.*
	10. Describe the statistical analysis.	*Yes. The statistical anlysis are described thorougly.*
Results	11. State the actual number of raters and subjects/objects which were included and the number of replicate observations which were conducted.	*Yes. This was described in the methods section.*
	12. Describe the sample characteristics of raters and subjects (e.g. training, experience).	*Yes. That is described.*
	13. Report estimates of reliability and agreement including measures of statistical uncertainty.	*Yes. That is described.*
Discussion	14. Discuss the practical relevance of results.	*Yes. The practical relevance is discussed.*
Auxiliary Material	15. Provide detailed results if possible (e.g. online)	*Yes. More detailed results available upon request.*

6.1.5 Comparing training sites

In Paper 3, we compared the performance between medical students at different training sites. As discussed, no NTS assessment tools can completely objectively and reliably measure NTS performance as long as humans are involved in the rating and interpretation (117). We considered scoring NTS performance using the NorMS-NTS assessment tool as the best alternative because, currently, no other options exist. Acknowledging the imperfections of the tool and its users, we factored these imperfections when conducting our research. We deemed it necessary for the same raters to rate all students in the study because the inter-rater reliability was not excellent. The number of videos and raters are further discussed in the *Sample Size* section.

As Norway and other countries decentralize medical education and because decentralization has been shown to increase the recruitment of physicians to rural areas, maintaining the highest quality of education is paramount. Factual medical knowledge, often referred to as technical skills or profession-specific competencies (3), can be compared using the final exam. The same examination was performed at all three training sites described in this thesis. The official numbers from UiT show a failure rate in the final year exam from 2018 to 2023 as follows: Tromsø 12.7% (out of 490 students), Bodø 8.8% (out of 147 students) and Finnmark 4% (out of 50 students). Notably, Bodø has a failure rate twice as high as that of Hammerfest, and Tromsø has a rate thrice as high as that of Hammerfest. This may be attributed to several factors. The number in Hammerfest comprises only 50 students, while there are 490 in Tromsø. Further research is necessary to confirm these results and determine the reasons for these differences. Nonetheless, this indicates that students in Hammerfest have high levels of technical skills.

A “good physician” requires technical skills and high NTS performance (121). In Paper 3, we found that Finnmark students also demonstrated higher levels of NTS performance than Bodø and Tromsø students. There was no significant difference between Tromsø and Bodø. These results support the validity of the NorMS-NTS tool. Tromsø and Bodø have largely the same learning activities and were expected to have similar NTS performance levels. However, Finnmark has different learning activities with a greater focus on NTS training; therefore, these differences are expected to be reflected in higher levels of NTS performance. This difference may be attributed to several factors, as discussed in Paper 3. Nevertheless, this suggests that NTS training provided in Finnmark is of good quality and effectively enhances NTS performance.

Overall, we argue that it is possible to provide high-quality, rural, and decentralized medical education. Medical students can be based in rural areas, undergo decentralized education, and attain elevated levels of NTS and technical proficiency. NTS and technical skills comprise the essential skills medical students require to transition from being students to becoming professionals and ultimately evolving into “good physicians” (121).

6.1.6 Patient safety

During my years of working on this thesis, the importance of patient safety in the project became increasingly clear to me. First, any mistake that harms a patient is devastating to the patient and their next of kin. However, mistakes that harm patients can also impact the responsible physician (10). Patient injuries also have repercussions for society. This entails additional societal costs, such as lost years of life, increased treatment costs, and workforce loss (11). Fifteen percent of the total cost of public hospitals in Norway is due to patient injuries (12). We aimed for the NorMS-NTS tool to make a difference, be easily implemented in medical schools, increase medical students’ NTS performance, and enhance new physicians' NTS performance, increasing patient safety. We envisioned NorMS-NTS as a tool that could train medical students in their ability to become lifelong professional learners.

Implementing assessment tools for NTS during medical school may raise students' awareness of NTS and enable future physicians to begin the process of being lifelong learners as professionals (54, 96). To maintain competence, physicians must be able to assess their practices (54). The healthcare system must facilitate this process. When introducing new standardized patient care pathways, strict guidelines, and procedures, healthcare professionals may lose the ability to develop flexibility, make decisions, handle uncertainty, and communicate with other health professionals and patients (85). Necessary guidelines are important, but they must be balanced toward a healthcare system that promotes teamwork, reflection, and critical thinking (80, 85). Physicians and other healthcare professionals can develop and continually improve NTS throughout their careers. Professionals are lifelong learners who should have work environments that give them opportunities to develop.

The NorMS-NTS and its implementation may be considered yet another guideline and tool that require time and effort. Hopefully, the need for minimal rating training and the fact that users find it useful will contribute to the successful implementation of this tool. Nevertheless, further studies are required to confirm this hypothesis.

6.2 Methodological considerations

6.2.1 Interviews

Participants in the focus group interviews in the first study were recruited through purposive sampling (116). We assumed that the people working with or treated by GP residents were experts whom NTS graduates should possess. Accordingly, we chose groups of professionals working with GP residents in emergency settings, including experienced general practitioners, paramedics, and ER nurses. As new interns started every six months and stayed for six months, the professionals interviewed worked with several new physicians. They observed new physicians from different universities with varying levels of NTS.

All groups were associated with the same hospital. All participants lived in the same county, and the entire area was rural. We considered the selection of groups to have varied sufficiently. The interviewed paramedics and ER nurses were from the same town. The physicians came from villages of various sizes near and far from the hospital. The distance from the hospital and number of inhabitants made the challenges different. We aimed to identify the geographical differences. If we had seen any geographic differences, we would have extended the interviews; however, the NTS mentioned were coherent and not associated with geographical location. The fact that we did not find any geographical differences underpins the possibility that the tool could be transferable to Norway as a whole, which should be confirmed in further studies.

The most important aspect of good NTS performance is patient benefit. We wanted to explore patients' views. Which NTS are the most important for the patient? What NTS do they think new physicians should hold? To explore this, we conducted interviews with the user committee at a local hospital, comprising patients and their next of kin of different ages and sexes. Committee members also hold roles in various patient-interest organizations. The members, including experienced patients and patients' relatives, live in different places; all bring different personal experiences to the table, mirroring the general population. We consider this group composition ideal for expanding and exploring knowledge about new physicians' NTS. Although we interviewed novices in the NTS field, they demonstrated a significant amount of implicit knowledge.

All interviews were conducted mono-professionally and were semi-structured (59). Hence, using the same vignette and interview guide in all interviews was highly important (122).

Upon reaching saturation, the same researcher conducted all subsequent interviews. After the first five interviews were completed, two more interviews were conducted to confirm the results (116). All interviews were recorded and transcribed.

All transcribed interviews were subsequently transferred to NVivo (AlfaSoft. com). We opted to conduct a modified open-card sort analysis because we did not want the researchers to be influenced by existing categories and began without preliminary categories (104). The modification was that the researchers sorted the data themselves instead of having others perform sorting. During the analysis process, a prototype of the tool was created and sent back to the participants to ensure that their opinions and inputs had been captured. We have received valuable feedback and made minor modifications to the tool.

6.2.2 Literature search

We conducted a literature search to review the comprehensive scientific work underlying existing tools and research. A literature search was conducted at the end of this process. We aimed for the process of creating a new tool to be as unaffected as possible by existing tools to determine whether new NTS would emerge. At this point in the process, we gathered all the inputs from the focus groups and created a prototype.

6.2.3 Sample size

The sample size in Paper 2 included 20 video-coded medical students, while in Paper 3, 24 video-coded medical students were included. The GRRAS analysis revealed a missing description of how the sample size was calculated in Paper 2. When calculating the sample size, it is necessary to know the underlying event rate (prevalence) and the standard deviation of the population, which are usually estimated from previous studies (123). Thus, the NorMS-NTS tool is completely new. We had no previous research on the tools or performance of the

Norwegian medical students' NTS. Therefore, the sample size was not calculated in this study.

Sample size calculations are challenging due to practical limitations. In Paper 2, three raters rated medical students in 20 videos, each approximately 15 minutes long (108). It was assumed that the balance between the number of videos each rater could rate and the number of videos needed for statistical analysis was met. The raters found it challenging to complete the ratings because they were time-consuming and required high levels of focus. Twenty videos were considered near the threshold for the number of videos each rater could rate. Based on this, we chose to compare eight students from each training site, with 24 videos for each rater. Ideally, several additional videos from each training site studied would be desirable; however, it would be impossible for the same rater to perform at high levels of focused ratings for so many videos. Therefore, as each rater rated only eight videos from each training site, we increased the number of raters required to rate each student to eight. Thus, the probability of accurately measuring NTS performance level increases as group size increases (124). Nonetheless, the sample size may be a limitation of the study and may limit its generalization. Further studies are warranted.

6.2.4 Reflexivity

In 2012, I graduated as a candidate from UiT — The Arctic University of Norway, and the last 10 years as a physician and general practitioner in Finnmark has influenced my preconceptions. We considered this when analyzing the data, with two separate researchers first performing the analysis individually. All validations and ratings of the tool were conducted by raters who were blinded to the training site and student identity; therefore, our previous work and analysis were validated by others.

During this period, the constructivist approach has become increasingly more prominent (125). We initially planned to validate the NorMS-NTS tool; nonetheless, we now know that the concept of validation is more nuanced than previously thought (105, 125). There is currently no validated NTS tool (105). A true and stable NTS score is not found if human factors are present (125). The assessment of the NTS remains an area with many unanswered questions.

6.2.5 User perspective

Higham et al.'s systematic review of observer-based tools for NTS assessment in healthcare described the development methods for different tools (60). None of the tools included a user perspective (patient representatives) as part of their development (60). In the development of the entire thesis project and creation of the NorMS-NTS tool, we found the patient's perspective to be fruitful and significant (103). NTS training aims to provide better healthcare services to patients. Hence, patients' perceptions should be considered. Interestingly, there was a high agreement on the necessary NTS between the focus groups comprising members of the user committees and the groups with health professionals. Desirable NTS seem universal for health professionals and patients. This underpins the validity of the NorMS-NTS tool and other similar existing tools.

The perspectives of the tool users are also important. If we want the tool to be relevant, implemented, and improve new physicians' NTS, it must be useful for users (98). In the second study, raters answered a questionnaire on the usability of the tool. This provides valuable information for further development and implementation of the NorMS-NTS tool. We consider our focus on user perspective a strength of the project and recommend the widespread use of the user perspective in research.

6.2.6 Open access

Open Access is important for the society to increase access to knowledge. In 2017, the Norwegian government aimed for all publicly funded scientific articles to be open access by the end of 2024 (126). The funders of this thesis, the Northern Norway Regional Health Authority, and the associated university UiT, require open-access publications (127, 128). The papers in this thesis are open-access. Papers 2 used open-access gold (108). Paper 1 presents green open access through the Munin Open Research Archive at UiT (103). This is a self-archiving system. As of January 1, 2022, UiT introduced its Right Retention Strategy to make all academic literature from UiT available through green Open Access (128). Thus, all articles were self-archived and accessible in Munin. Articles published in gold open-access journals are deposited as publishers' PDF. Articles from the closed subscription journal UiT have been deposited as the latest peer-reviewed version of the manuscript. UiT rector has legal responsibility for this policy.

The NorMS-NTS tool and handbook are freely available from the Finnish Hospital Trust website (129). All relevant research is freely accessible through Open Access. We consider this an advantage of the widespread use of the tool.

7 Conclusions

We have developed an assessment tool for Norwegian medical students' NTS, the NorMS-NTS. The process of collecting evidence of validity have started. We found that the tool was usable for raters. Interrater reliability, internal consistency, and observability were sufficient for formative assessments when used by novice raters. We observed that rural, decentralized medical students in Finnmark had higher levels of NTS performance than the students in Bodø and Tromsø.

In rural areas, decentralized medical education may contribute to the development of highly skilled physicians. Medical students transition from being students to becoming highly qualified professionals, encompassing NTS and factual medical knowledge, as a result of their rural, decentralized education.

8 Future perspectives

Our main goal was to create a tool to assess Norwegian medical students' NTS and compare their performance at different training sites. During the course of this thesis, the necessity of high NTS performance for patient safety became clear. This thesis is one step toward higher NTS levels for new physicians in Norway. The NorMS-NTS tool is only the beginning, fulfilling national expectations to perform higher-quality NTS training in medical schools. This offers the possibility of improving new physicians' NTS and patient safety. To do this, it is necessary to evaluate the training given and see whether the students amend their NTS performance.

If the NorMS-NTS tool ends with this thesis, it will not change anything. Further research is required in this area. We aspire to see this tool implemented in Norwegian medical schools as a formative tool for student feedback, enhancing awareness of the importance of NTS. To accomplish this, it is necessary to fully validate the formative assessment tool by further examining its consequences. This involves determining whether the assessments obtained by the tool are correct and beneficial, as well as assessing the tool's impact on students.

Further research should focus on the usability of this tool for summative assessment. The tool can then be used to evaluate the education provided to improve NTS training. The tool should also be validated for Norwegian physicians in the first part of their specialization (LIS1).

Furthermore, the tool should be validated in different settings and situations, including other Scandinavian countries.

If the tool is implemented, the effects of its widespread use should be studied. Will structured NTS performance training, evaluation, and feedback during medical school influence physicians' clinical work? Based on the currently available research, we believe that the implementation of NorMS-NTS may contribute to better teaching and feedback and provide medical students in Norway with better NTS (45). In such cases, newly qualified physicians will eventually have better NTS. This may reduce the risk of adverse events and improve patient safety. Nonetheless, further studies are required to confirm this hypothesis.

Figure 9 Hammerfest - Finnmark. Photo: Katrine Prydz

References

1. Smith CM. Origin and uses of primum non nocere - above all, do no harm. J Clin Pharmacol. 2005;45(4):371-7.
2. Lawton R, Johnson J, Janes G, Foy R, Simms-Ellis R. Supporting doctors who make mistakes. BMJ. 2019;365:l2161.
3. Donaldson L, Ricciardi W, Sheridan S, Tartaglia R. Textbook of Patient Safety and Clinical Risk Management. Cham: Springer International Publishing : Imprint: Springer; 2021.
4. Hays RB, Lawson M, Gray C. Problems presented by medical students seeking support: A possible intervention framework. Med Teach. 2011;33(2):161-4.
5. Weurlander M. Becoming a physician involves learning to manage uncertainty and learning how to fail. Med Educ. 2020;54(9):776-8.
6. Nevalainen MK, Mantyranta T, Pitkala KH. Facing uncertainty as a medical student—A qualitative study of their reflective learning diaries and writings on specific themes during the first clinical year. Patient Educ Couns. 2010;78(2):218-23.
7. Frank JR, Danoff D. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. Med Teach. 2007;29(7):642-7.
8. Malling B, Eika B. The CanMEDS' professional role is complex and needs redefinition in Denmark. Ugeskr Laeger. 2013;175(24):1709-12.
9. Warren AE, Allen VM, Bergin F, Hazelton L, Alexiadis-Brown P, Lightfoot K, et al. Understanding, teaching and assessing the elements of the CanMEDS professional role: Canadian Program Directors' views. Med Teach. 2014;36(5):390-402.
10. Epstein RM, Hundert EM. Defining and assessing professional competence. Jama. 2002;287(2):226-35.
11. Epstein R. Assessment in Medical Education. N Engl J Med. 2007;356(4):387-96.
12. Leach DC. Competence is a habit. Jama. 2002;287(2):243-4.
13. Fletcher G, McGeorge P, Flin R, RJ G, Maran N. The role of non-technical skills in anaesthesia: a review of current literature. British journal of anaesthesia. 2002;88(3):418-29.
14. Kunnskapsdepartementet. Forskrift om nasjonal retningslinje for medisinutdanning. 2020.
15. Bell CL, Allan JL, Ross S, Powell DJH, Johnston DW. How can we better prepare new doctors for the tasks and challenges of ward rounds?: An observational study of junior doctors' experiences. Med Teach. 2021;43(11):1294-301.
16. Goodyear HM. First year doctors experience of work related wellbeing and implications for educational provision. Int J Med Educ. 2014;5:103-9.
17. UiT F. The Finnmark Model 2016 [Available from: https://uit.no/Content/473310/PROSJEKTRAPPORT-FASE-1-FINNMARKSMODELLEN-med-vedlegg.pdf.
18. Royal College of Physicians and Surgeons of Canada. CanMEDS role: Professional. [Available from: http://www.royalcollege.ca/rcsite/canmeds/framework/canmeds-role-professional-e.
19. Duffy FD, Lynn LA, Didura H, Hess B, Caverzagie K, Grosso L, et al. Self-assessment of practice performance: development of the ABIM Practice Improvement Module (PIM). J Contin Educ Health Prof. 2008;28(1):38-46.

20. Murphy P, Nestel D, Gormley GJ. Words matter: towards a new lexicon for 'nontechnical skills' training. Adv Simul. 2019;4(1):8.
21. Nestel D, Walker K, Simon R, Aggarwal R, Andreatta P. Nontechnical skills: An inaccurate and unhelpful descriptor? Simul Healthc. 2011;6(1):2-3.
22. Flin RH, Crichton M, O'Connor P. Safety at the sharp end : a guide to non-technical skills. Boca Raton, Fla.: CRC Press; 2008.
23. Flin R OCP. Safety at the Sharp End: A Guide to Non-technical Skills: Ashgate; 2008.
24. Prineas S, Mosier K, Mirko C, Guicciardi S. Non-technical skills in healthcare. In: Donaldson L, Ricciardi W, Sheridan S, Tartaglia R, editors. Textbook of Patient Safety and Clinical Risk Management. Cham: Springer International Publishing; 2021. p. 413-34.
25. Flin R, Patey R. Improving patient safety through training in non-technical skills. BMJ. 2009;339:b3595.
26. Kanki BG, Helmreich RL, Anca JM. Crew Resource Management. Amsterdam, Boston: Academic Press/Elsevier; 2010.
27. Ford J, Henderson R, O'Hare D. The effects of crew resource management (CRM) training on flight attendants' safety attitudes. J Safety Res. 2014;48:49-56.
28. Metscher DS, Smith M, Alghamdi A. Multi-cultural factors in the crew resource management environment: promoting aviation safety for airline operations. JAAER. 2009;18(2):9-23.
29. DV D, J KH, AL A. Human factors in the investigation of accidents and incidents. Eur Air Traffic Control Harmon Integr Program HUM ET1 ST131998.
30. A S, K T-e. Exploring technical and non-technical competencies of navigators for autonomous shipping2021:[1-19 pp.].
31. Owen C. Human Factors Challenges in Emergency Management: Enhancing Individual and Team Performance in Fire and Emergency Services. 1st edition. ed: United States: CRC Press; 2014.
32. Havinga J, de Boer RJ, Rae A, Dekker S. How did crew resource management take-off outside of the cockpit? A systematic review of how crew resource management training is conceptualised and evaluated for non-pilots. Safety. 2017;3(4):26.
33. Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. Shifting paradigms: From Flexner to competencies. Acad Med. 2002;77(5):361-7.
34. Frank J, Mungroo R, Ahmad Y, M W, De Rossi S, Horsley T. Toward a definition of competency-based education in medicine: a systematic review of published definitions. Med Teach. 2010;32(8):631-7.
35. Ferguson PC, Caverzagie KJ, Nousiainen MT, Snell L. Changing the culture of medical training: An important step toward the implementation of competency-based medical education. Med Teach. 2017;39(6):599-602.
36. Hamza DM, Hauer KE, Oswald A, van Melle E, Ladak Z, Zuna I, et al. Making sense of competency-based medical education (CBME) literary conversations: A BEME scoping review: BEME Guide No. 78. Med Teach. 2023;45(8):802-15.
37. Health Comitee HoC UK. Health Comitee - Sixth Report, Patient Safety. 2009.
38. Stergiopoulos V, Lieff S, Razack S, al. e. Canadian residents' perceived manager training needs. Med Teach. 2010;32(11):e479-85.
39. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. Lancet. 2010;376(9756):1923-58.
40. Grimstadutvalget. Studieplasser i medisin i Norge. Behov, modeller og muligheter. 2019.

41. Hagemann V, Herbstreit F, Kehren C, Chittamadathil J, Wolfertz S, Dirkmann D, et al. Does teaching non-technical skills to medical students improve those skills and simulated patient outcome? Int J Med Educ. 2017;8:101-13.
42. Somasundram K, Spence H, Colquhoun AJ, McIlhenny C, Biyani CS, Jain S. Simulation in urology to train non-technical skills in ward rounds. 2018:705-12.
43. Yeung J, Husselbee N, Davies R, Melody T, Lockey A, Gwinutt C, et al. Effect of non-technical skills teaching on performance – An evaluation of additional non-technical skills teaching in Resuscitation Council ALS course. Resuscitation. 2015;96:24-.
44. Nicolaides M, Theodorou E, Hanrahan JG, Theodoulou I, Emin EI, Papalois A, et al. Advancing medical students’ non-technical skills in a group-based setting. J Invest Surg. 2021;34(1):39-43.
45. Gordon M, Darbyshire D, Baker P. Non-technical skills training to enhance patient safety: a systematic review. Med Educ. 2012;46(11):1042-54.
46. Dieckmann P, Glavin R, Hartvigsen Grønholm Jepsen RM, Krage R. Non-technical skills bingo—a game to facilitate the learning of complex concepts. Adv Simul. 2016;1(1):23.
47. Jepsen RMHG, Østergaard D, Dieckmann P. Development of instruments for assessment of individuals’ and teams’ non-technical skills in healthcare: a critical review. Cognition, technology & work. 2015;17(1):63-77.
48. Zajac S, Woods A, Tannenbaum S, Salas E, Holladay CL. Overcoming challenges to teamwork in healthcare: A team effectiveness framework and evidence-based guidance. Front Commun. 2021;6.
49. Stephens T, Hunningher A, Mills H, Freeth D. An interprofessional training course in crises and human factors for perioperative teams. J Interprof Care. 2016;30(5):685-8.
50. NOU 2015:11 Med åpne kort. Norwegian Official Report. 2015.
51. – Dette var utfordrende, morsomt, lærerikt og på høy tid å få gjennomført [Available from: https://uit.no/nyheter/artikkel?p_document_id=411077&p_dim=88128.
52. Øver på samarbeid for å unngå feilbehandling [Available from: https://uit.no/nyheter/artikkel?p_document_id=604327.
53. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. Adv Health Sci Educ. 2014;19(2):233-50.
54. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. Med Teac. 2010;32(8):676-82.
55. Lyk-Jensen HT, Jepsen RM, Spanager L, Dieckmann P, Østergaard D. Assessing nurse anaesthetists' non-technical skills in the operating room. Acta anaesthesiologica Scandinavica. 2014;58(7):794-801.
56. Jepsen RM, Spanager L, Lyk-Jensen HT, et.al. Customisation of an instrument to assess anaesthesiologists' non-technical skills. Int J Med Educ. 2015;6:17-25.
57. Spanager L, Lyk-Jensen H, Dieckmann P, Wettergren A, Rosenberg J, Østergaard D. Customization of a tool to assess Danish surgeons’ non-technical skills in the operating room. Dan Med J. 2012(59(11)):A4526.
58. Hamilton AL, Kerins J, MacCrossan MA, Tallentire VR. Medical students’ non-technical skills (Medi-StuNTS): Preliminary work developing a behavioural marker system for the non-technical skills of medical students in acute care. BMJ Simul Technol Enhanc Learn. 2019;5(3):130-9.
59. Walker S, Brett S, McKay A, Lambden S, Vincent C, Sevdalis N. Observational skill-based clinical assessment tool for resuscitation (OSCAR): Development and validation. Resuscitation. 2011;82(7):835-44.

60. Higham H, Greig PR, Rutherford J, Vincent L, Young D, Vincent C. Observer-based tools for non-technical skills assessment in simulated and real clinical environments in healthcare: a systematic review. BMJ Qual Saf. 2019;28(8):672-86.
61. Cooper S, Cant R, Porter J, Sellick K, Somers G, Kinsman L, et al. Rating medical emergency teamwork performance: Development of the team emergency assessment measure (TEAM). Resuscitation. 2010;81(4):446-52.
62. Malec JF, Torsher LC, Dunn WF, Wiegmann DA, Arnold JJ, Brown DA, et al. The mayo high performance teamwork scale: reliability and validity for evaluating key crew resource management skills. Simulation in healthcare : journal of the Society for Simulation in Healthcare. 2007;2(1):4-10.
63. Yule S, Flin R, Maran N, Rowley D, Youngson G, Paterson-Brown S. Surgeons' non-technical skills in the operating room: Reliability testing of the NOTSS behavior rating system. World J Surg. 2008;32(4):548-56.
64. Russ S, Hull L, Rout S, Vincent C, Darzi A, Sevdalis N. Observational teamwork assessment for surgery: Feasibility of clinical and nonclinical assessor calibration with short-term training. Ann Surg. 2012;255(4):804-9.
65. Higham H, Greig P, Crabtree N, Hadjipavlou G, Young D, Vincent C. A study of validity and usability evidence for non-technical skills assessment tools in simulated adult resuscitation scenarios. BMC Med Educ. 2023;23(1):153.
66. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess. 1994;6(4):284-90.
67. Hull L, Arora S, Symons NR, et.al. Training faculty in nontechnical skill assessment: national guidelines on program requirements. Ann Surg. 2013;258(2):370-5.
68. Helsedirektoratet. Pasientskader i Norge 2022 - Målt med Global Trigger Tool. 2023.
69. Services NMoHaC. High Quality - Safe Services. Quality and Patient Safety in the Healh and Care Services. Meld. St. 10 (2012–2013) Report to the Storting (white paper) Summary. 2012-2013.
70. Helsedirektoratet. Nasjonal handlingsplan for pasientsikkerhet og kvalitetsforbedring 2019-2023. 2018.
71. Larsen IMF. Veien mot Helseplattformen Bricolage som døråpner for større innovasjonsprosjekt: NTNU; 2019.
72. Vardenær AMS. Sosial ulikhet i helse hos eldre. En diskursanalyse av hvordan eldre posisjoneres i Meld. St. 15 (2017-2018) - Leve hele livet: UiT Norges arktiske universitet; 2022.
73. Helsedirektoratet. Ledelse og kvalitetsforbedring i helse- og omsorgstjenesten. 2017.
74. Helsedirektoratet. Pasientskader i Norge 2021. 2022.
75. The Norwegian System of Patient Injury Compensation [Available from: www.npe.no.
76. The World Alliance of Patient Safety [Available from: https://www.who.int/teams/integrated-health-services/patient-safety/about/world-alliance-for-patient-safety.
77. Vincent C, Aylin P, Franklin BD, Holmes A, Iskander S, Jacklin A, et al. Is health care getting safer? Bmj. 2008;337:a2426.
78. Dekker SWA. The bureaucratization of safety. Saf sci. 2014;70:348-57.
79. Helsedirektoratet. Handlingsplan mot antibiotikaresistens i helsetjenesten - evalueringsrapport 2022. 2023.
80. Provan DJ, Dekker SWA, Rae AJ. Bureaucracy, influence and beliefs: A literature review of the factors shaping the role of a safety professional. Saf Sci. 2017;98:98-112.
81. Porter J, Boyd C, Skandari MR, Laiteerapong N. Revisiting the time needed to provide adult primary care. J Gen Intern Med. 2023;38(1):147-55.

82. Institute of M, National Academy of E, Fanjiang G, Grossman JH, Compton WD, Reid PP. Building a Better Delivery System : A New Engineering/Health Care Partnership. Washington, D.C., UNITED STATES: National Academies Press; 2005.
83. Carayon P. Human factors in patient safety as an innovation. Appl Ergon. 2010;41(5):657-65.
84. Maurino DE, Salas E. Human Factors in Aviation. Amsterdam, Boston: Academic Press/Elsevier; 2010.
85. Dekker S. Patient Safety : A Human Factors Approach. Baton Rouge, United States: Taylor & Francis Group; 2011.
86. Slawomirski L, Auraaen A, Klazinga NS. The economics of patient safety: Strengthening a value-based approach to reducing patient harm at national level. OECD Health Working Papers 96, OECD Publishing; . 2017.
87. Mlambo M, Dreyer A, Dube R, Mapukata N, Couper I, Cooke R. Transformation of medical education through Decentralised Training Platforms a scoping review. Rural Remote Health. 2018;18(1):1-16.
88. Boelen C, Woollard B. Social accountability and accreditation: A new frontier for educational institutions. Med Educ. 2009;43(9):887-94.
89. Kiguli-Malwadde E, Olapade-Olaopa EO, Kiguli S, Chen C, Sewankambo NK, Ogunniyi AO, et al. Competency-based medical education in two Sub-Saharan African medical schools. Adv Med Educ Pract. 2014;5:483-9.
90. Strasser R, Neusy AJ. Context counts: Training health workers in and for rural and remote areas. Bull World Health Organ. 2010;88(10):777-82.
91. Westlie Å L, Gaski M, Abelsen B, Grimstad H, Ofstad EH. Doctors educated in Bodø - who are they and where do they work? Tidsskr Nor Laegeforen. 2022;142(1).
92. Nilsen G, Huemer J, Eriksen L. Bachelor studies for nurses organised in rural contexts – a tool for improving the health care services in circumpolar region? Int J Circumpolar Health. 2012;71(1):17902.
93. Eriksen LT, Huemer JE. The contribution of decentralised nursing education to social responsibility in rural Arctic Norway. Int J Circumpolar Health. 2019;78(1):1691706.
94. Mezirow J. Transformative Dimensions of Adult Learning: San Francisco, CA: Jossey Bass; 1991.
95. Mezirow J. Transformative learning as discourse. JTransform Educ. 2003;1(1):58-63.
96. Kerins J, Smith SE, Phillips EC, Clarke B, Hamilton AL, Tallentire VR. Exploring transformative learning when developing medical students' non-technical skills. Med Educ. 2020;54(3):264-74.
97. User participation in Health Research in Norway -Final guidelines and measures 26/02/2015 2015 [Available from: https://www.helse-nord.no/499f62/siteassets/dokumenter-og-blokker/forskning/forskningsmidler/brukermedvirkning-i-helseforskning-i-norge-vedtatte-retningslinjer-og-tiltak-til-rhf-styresak-19-2015.pdf.
98. Guide for User Participation in Health Research in the Specialist Health Service 2018 [Available from: https://www.helse-nord.no/499f71/siteassets/dokumenter-og-blokker/forskning/forskningsmidler/2020-forskningsmidler/veileder-brukermedvirkning-i-forskning-mai-2018.pdf.
99. Olsen IP, Lillebo B, Ofstad EH, Østlyngen A, Lappegård KT. Time to expand decentralisation of medical education in Norway. Tidsskr Norske Lægeforening. 2021;141(11).
100. Ringsted C, Hodges B, Scherpbier A. 'The research compass': an introduction to research in medical education: AMEE Guide no. 56. Med Teach. 2011;33(9):695-709.
101. Johnson RB, Onwuegbuzie AJ, Turner LA. Toward a definition of mixed methods research. J Mix Methods Res. 2007;1(2):112-33.

102. Schoonenboom J, Johnson RB. How to construct a mixed methods research design. Kolner Z Soz Sozpsychol. 2017;69(Suppl 2):107-31.
103. Prydz K, Dieckmann P, Musson D, Wisborg T. The development of a tool to assess medical students' non-technical skills – The Norwegian medical students' non-technical skills (NorMS-NTS). Med Teach. 2022:1-8.
104. Fincher S, Tenenberg J. Making sense of card sorting data. Expert Syst. 2005;22(3):89-93.
105. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. Adv Simul. 2016;1(1):31.
106. Association AER. American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association; 2014.
107. Hill J, Ogle K, Gottlieb M, Santen SA, Artino AR, Jr. Educator's blueprint: A how-to guide for collecting validity evidence in survey- based research. AEM Educ Train. 2022;6(6):e10835.
108. Prydz K, Dieckmann P, Fagertun H, Musson D, Wisborg T. Collecting evidence of validity for an assessment tool for Norwegian medical students' non-technical skills (NorMS-NTS): usability and reliability when used by novice raters. BMC Med Educ. 2023;23(1):865.
109. Developing and Validating Rapid Assessment Instruments. Washington: Washington: Taylor & Francis Ltd; 2010. p. 158.
110. Montgomery DC. Design and analysis of experiments. 8th ed. Hoboken, N.J: Wiley; 2013.
111. Stoll A. Post hoc tests: Tukey honestly significant difference test. The SAGE Encyclopedia of Communication Research Methods 4: SAGE Publications, Inc; 2017.
112. Frithioff A, Frendø M, Foghsgaard S, Sørensen MS, Andersen SAW. Are video recordings reliable for assessing surgical performance? A prospective reliability study using generalizability theory. Simulation in healthcare : journal of the Society for Simulation in Healthcare. 2023;18(4):219-25.
113. King HB, Battles J, Baker DP, Alonso A, Salas E, Webster J, et al. Advances in Patient Safety TeamSTEPPS(™): Team strategies and tools to enhance performance and patient safety. In: Henriksen K, Battles JB, Keyes MA, Grady ML, editors. Advances in Patient Safety: New Directions and Alternative Approaches (Vol 3: Performance and Tools). Rockville (MD): Agency for Healthcare Research and Quality (US); 2008.
114. Endsley MR, Garland DJ. Situation awareness : analysis and measurement. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.
115. Salas E, DiazGranados D, Klein C, Burke CS, Stagl KC, Goodwin GF, et al. Does team training improve team performance? A meta-analysis. Hum Factors. 2008;50(6):903-33.
116. Malterud K. Kvalitative forskningsmetoder for medisin og helsefag. 4 ed: Universitetsforlaget; 2017.
117. Mommers L, Verstegen D, Dolmans D, van Mook WNKA. Observation of behavioural skills by medical simulation facilitators: a cross-sectional analysis of self-reported importance, difficulties, observation strategies and expertise development. Adv Simul. 2023;8(1):28-.
118. LeBreton JM, Senter JL. Answers to 20 questions about interrater reliability and interrater agreement. Organ Res Methods. 2008;11(4):815-52.
119. Clarke B, Smith SE, Phillips EC, Hamilton A, Kerins J, Tallentire VR. Reliability of assessment of medical students' non-technical skills using a behavioural marker system: does clinical experience matter? BMJ Simul Technol Enhanc Learn. 2021;7(5):285-92.

120. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64(1):96-106.
121. Steiner-Hofbauer V, Schrank B, Holzinger A. What is a good doctor? Wien Med Wochenschr. 2018;168(15):398-405.
122. Malterud K. Fokusgruppper som forskningsmetode for medisin og helsefag: Universitetsforlaget; 2012.
123. Kadam P, Bhalerao S. Sample size calculation. Int J Ayurveda Res. 2010;1(1):55-7.
124. Andrade C. Sample size and its importance in research. Indian J Psychol Med. 2020;42(1):102-3.
125. Schuwirth L, Colliver J, Gruppen L, et.al. Research in assessment: consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach. 2011;33(3):224-33.
126. Research MoEa. National goals and guidelines for open access to research articles. In: Affairs DoL, editor. 2017.
127. Authority TNNRH. The Northern Norway Regional Health Authorities' publication fund for researchers 2022 [Available from: https://www.helse-nord.no/forskning-og-innovasjon/helse-nords-publiseringsfond-for-forskere.
128. Principles for open access to academic publications at UiT The Arctic University of Norway, (2021).
129. Trust FHH. NorMS-NTS Handbook 2023 [Available from: https://www.finnmarkssykehuset.no/fag-og-forskning/lis/norms-nts-handbok/.

Paper 1

Paper 2

Prydz *et al. BMC Medical Education* (2023) 23:865
https://doi.org/10.1186/s12909-023-04837-6

BMC Medical Education

RESEARCH

Open Access

Collecting evidence of validity for an assessment tool for Norwegian medical students' non-technical skills (NorMS-NTS): usability and reliability when used by novice raters

Katrine Prydz[1,2*], Peter Dieckmann[3,4,5], Hans Fagertun[6], David Musson[7] and Torben Wisborg[1,2]

Abstract

Background The NorMS-NTS tool is an assessment tool for assessing Norwegian medical students' non-technical skills (NTS). The NorMS-NTS was designed to provide student feedback, training evaluations, and skill-level comparisons among students at different study sites. Rather than requiring extensive rater training, the tool should capably suit the needs of busy doctors as near-peer educators. The aim of this study was to examine the usability and preliminary assess validity of the NorMS-NTS tool when used by novice raters.

Methods This study focused on the usability of the assessment tool and its internal structure. Three raters used the NorMS-NTS tool to individually rate the team leader, a medical student, in 20 video-recorded multi-professional simulation-based team trainings. Based on these ratings, we examined the tools' internal structure by calculating the intraclass correlation coefficient (ICC) (version 3.1) interrater reliability, internal consistency, and observability. After the rating process was completed, the raters answered a questionnaire about the tool's usability.

Results The ICC agreement and the sum of the overall global scores for all raters were fair: ICC (3,1) = 0.53. The correlation coefficients for the pooled raters were in the range of 0.77–0.91. Cronbach's alpha for elements, categories and global score were mostly above 0.90. The observability was high (95%-100%). All the raters found the tool easy to use, none of the elements were redundant, and the written instructions were helpful. The raters also found the tool easier to use once they had acclimated to it. All the raters stated that they could use the tool for both training and teaching.

Conclusions The observed ICC agreement was 0.08 below the suggested ICC level for formative assessment (above 0.60). However, we know that the suggestion is based on the average ICC, which is always higher than a single-measure ICC. There are currently no suggested levels for single-measure ICC, but other validated NTS tools have single-measure ICC in the same range. We consider NorMS-NTS as a usable tool for formative assessment of Norwegian medical students' non-technical skills during multi-professional team training by raters who are new to the tool. It is necessary to further examine validity and the consequences of the tool to fully validate it for formative assessments.

*Correspondence:
Katrine Prydz
katrine.prydz@uit.no
Full list of author information is available at the end of the article

Keywords NorMS-NTS, Nontechnical skills, Medical students, Assessment, Simulation-based training, Validation, Assessment tools

Background

Non-technical skills (NTSs) are defined as 'the cognitive, social and personal resource skills that complement technical skills and contribute to safe and efficient task performance' [1]. Examples of NTSs include skills in decision making, leadership, teamwork, situation awareness, etc. [2]. Studies show that NTSs can be improved through training [3–6]. Medical students need to learn NTSs during medical school, as the high-level use of NTSs is important for patient safety [1, 7]. Poor NTS performance has been identified as a contributing factor in 70% of the adverse events that occur in hospitals [8].

Training NTS requires an NTS assessment tool to ensure that medical students successfully obtain these skills during medical school. NTS tools can be used to evaluate students' NTS performance, give them feedback and evaluate the NTS training. Several tools have been developed for the assessment of health professionals' NTSs [9–14]. The most versatile and flexible is the Scottish Anesthetists Non-Technical Skills rating system (ANTS) [9]. This has been further developed into Danish and Norwegian adaptations aimed at assessing nurse anesthetists [15]. Other tools are the Non-Technical Skills for Surgeons (NOTSS) [16], Anesthetists Non-Technical skills for Anesthesia Practitioners (ANTS-AP) [17] and the Scrub Practitioners' List of Intraoperative Non-Technical Skills (SPLINTS) [18]. For medical students, the Medical Students' Non-Technical Skills (Medi-StuNTS) [19] was created in the United Kingdom [20]. There is also a tool for anesthesiology students, the Anesthesiology Students' Non-Technical skills (AS-NTS) [14].

There is evidence of the need to develop customized tools for each profession and even for specific countries and cultures [21, 22]. Different countries have differences in culture, tasks and responsibilities, which likely require contextualizing what NTS is about and how they would be used. Studies have found that NTS tools developed in the United Kingdom had to be adapted for use in a Danish setting [22]. To avoid a risk of overlooking specific desired NTS for Norwegian Medical students if adapting an existing tool. we decided to create a new tool to assess Norwegian medical students' nontechnical skills (NorMS-NTS) [23]. The process of the development of NorMS-NTS has been thoroughly described previously [23].

NorMS-NTS was created as a tool for assessing NTS in relation to student feedback, training evaluations, and comparing student skills levels among different study sites. To facilitate a broader adoption of the tool and to optimize the validation of data, the ease of use was a critical feature for this tool. That the tool does not require extensive rater training was thus of importance.

The aim of this study was to examine the usability and preliminary assess validity of the NorMS-NTS tool when used by novice raters.

We recognize that validity interpretation is not simply a matter of either being valid or not [24]. The issue of validity is measured through scores, interpretation, and use, not simply by the tool. Different uses of the same tool may lead to diverging results. In other words, validity is context dependent. When validating NTS assessment tools, it is important to define and clearly specify the intended context. Evidence validated in one specific setting is often transferable to another setting, but that should be specifically determined according to each situation. Validation is a continuous process of collecting evidence over time and in different contexts.

As the aim of this first part of the validation process was to examine novices' use of the NorMS-NTS. Our focus in this study is the usability of the tool and its internal structure, as measured by interrater reliability, internal consistency, and observability. A full validation for formative assessment with consequences and impact on students is beyond the scope of this article. We did not collect validity evidence for the use of the tool for summative assessment, as it requires extended rater training. Previous studies from aviation show that even those who know human factors need 2–3 days of training and calibration to reach sufficient single rater inter-rater reliability [25].

Methods

The NorMS-NTS consists of four categories, 13 elements and an overall score (Table 1). The categories and elements are rated on a 5-point Likert scale, and the overall global scores are rated on a 7-point Likert scale.

Validity evidence was collected by performing as an observational study using three raters to assess the human performance evidenced in 20 videos. Three doctors from RegSim were recruited as raters. RegSim is a unit at the Northern Norway Regional Health Authority (Helse Nord) that is responsible for simulation training in all hospitals in northern Norway. All three doctors had broad clinical experience and shared a stated interest in simulation (Table 2). The raters were blinded to the participants' educational grade. The three raters were required to read the NorMS-NTS training manual

Table 1 NorMS-NTS

Category[a]	Category score[b]	Element[a]	Element score[b]	Feedback
Communication		Team communication		
		Establish mutual understanding		
		Patient communication		
Situation awareness		Situational assessment		
		Understanding of team members' roles		
		Attentiveness		
Teamwork		Professional modesty		
		Flexibility		
		Efficient use of team members		
Decision making		Uncertainty management		
		Decision analysis		
		Leadership		
		Prioritization		

General comments: ____

[a] N/A – Not applicable. 1, much below average; 2, below average; 3, acceptable; 4, above average; 5, much above average

[b] Within team unless otherwise specified

Overall global rating (marked with a ring):

Very poor 1–2 – 3–4 – 5–6 – 7 Excellent

Table 2 Raters' backgrounds

Background	Rater 1	Rater 2	Rater 3
Age	*57*	*51*	*46*
Specialization	*Pediatrician*	*Anesthesiologist*	*Anesthesiologist*
Academic competency highest degree/position?	*PhD*	*Cand. med*	*Cand. med*
Clinical experience (number of years in clinical practice)	*30 years*	*25 years*	*19 years*
Do you have any prior experience with nontechnical skills (NTSs) or tools for NTS assessment?	*No*	*Yes*	*Yes, many years of experience with simulation training, but not with specific tools like this*

developed by the author (KP). The research team member KP delivered a 20-min overview of the tool to all three raters via Microsoft Teams ®. The three raters were then given online access to the videos through an online data portal. Raters received the tool through e-mail. Each rater individually rated the team leader (medical student) through 20 video-recorded multiprofessional simulation-based team trainings using the NorMS-NTS tool. One rater completed the forms electronically and sent them to researcher KP via email. The remaining two raters printed the forms and filled them out manually, then they scanned them and returned them via e-mail.

Each video was assigned a study identification number consisting of two digits, and the three raters were assigned the numbers 01, 02 or 03. The data from the raters' marking sheets were entered into an Excel sheet. The data were then imported into the Statistical Analysis System (SAS© ver. 9.4) for analysis. The data were checked for possible errors, such as incorrect scales or missing ratings. Then, the data were stored in a permanent and password-protected SAS database in preparation for the analyses.

Setting

The medical students participating in this study were enrolled as students at UiT—The Arctic University of Norway in Hammerfest, Tromsø and Bodø. All students had multi-professional team training as part of their curriculum. The teams mostly consisted of medical students and nursing students, although some teams also had radiography students or bioengineering students on their team. The medical students were in their 5th and 6th years of study. Two different simulation-based training scenarios were used, and they were implemented following detailed descriptions in scenario scripts. Each simulation lasted between 12–20 min. Due to COVID-19 restrictions, some of the scenarios were implemented using a simulation manikin rather than a simulated patient. All scenarios had a trained nurse or doctor as the facilitator. The simulated patient was examined, answered student questions, and expressed pain and emotions. The students performed all measures and examinations, and the facilitator then informed them of the results consecutively. If the desired equipment was not available, the students were told to say what they would have done, which is a low-cost, easily accessible method of simulation training that can be performed anywhere.

Ethics

Norwegian law exempts educational studies from ethical approval because they do not involve patients. However, the Regional Committee of North Norway for Medical and Health Research provided feedback on the protocol used in this study and approved this assumption (Ref: 2016/1539/REK nord). The participant consent form was approved by the Norwegian Center of Research Data (Ref: 57,474/2017). Informed consent from all participants was obtained after oral and written information was delivered on the purpose and objectives of the study.

The rating of the videos was performed on the Services for sensitive data (TSD) facilities owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT department (USIT). All videos were saved at the TSD. TSD provides a platform for public research institutions in Norway. This service provides a secure project area where researchers can collect, store, and analyze sensitive data.

Validity dimensions

Messick's framework is recommended as a method of collecting evidence to validate assessment tools [24]. There are other frameworks available, but we chose Messick's, as it has been the standard in the field since 1999 [26]. It is a conceptual, theoretical framework that utilizes five sources of evidence: content, internal structure, relationship with other variables, response process and consequences. We have summarized our validation procedures for different sources in Table 3, which displays the different dimensions we used to investigate validity evidence regarding the use of the NorMS-NTS.

Content

Evidence for validation of the tool's content was collected during the development of the NorMS-NTS [23]. The tool was created based on information gathered from focus group interviews. Participants in these focus groups provided their views regarding which NTS were necessary for newly graduated physicians. After analyzing the interviews, the participants were asked to provide feedback regarding the tool. Participants were asked if the tool accurately reflected their opinions and inputs. The feedback provided indicated that the assessment tool accurately reflected their opinions. Despite beginning the tool's development from scratch, the tool was quite similar to previously described tools, demonstrating convergent validity and thus supporting content validity [9, 21, 28, 29].

Internal structure

Interrater reliability

ICC (3,1) was calculated as all subjects were being rated by the same specific population of raters. The nonparametric statistic Kendall's W was also used to assess the level of agreement between raters.

Internal consistency analysis

The correlation between the elements, categories and overall global scores was measured. The Spearman nonparametric correlation between each category and the corresponding elements was calculated, as well as that between the global scores and the categories. In addition, Cronbach's alpha (CA) was applied.

Table 3 Messick framework: sources of evidence, definitions and procedure

Source of evidence	Definition	Procedure
Content	"the relationship between a test's content and the construct it is intended to measure [26]."	*Assessed as a part of development*
Internal structure	"The relationship among data items within the assessment and how these relate to the overarching construct [24]"	Interrater reliability Internal consistency Observability
Relationships with other variables	"The degree to which these relationships are consistent with the construct underlying the proposed test score interpretations [26]"	*Planned in further validations*
Response process	"The fit between the construct and the detailed nature of performance... actually engaged in [26]"	Raters respond in questionnaire
Consequences	"The impact, beneficial or harmful and intended or unintended, of assessment [27]"	Evaluation of the possibility of minimal rater training

Observability

The observability of each element, category and global score was calculated by the percentage of observations recorded by the raters. An observability > 50% is deemed acceptable [30].

Response process

All raters received a questionnaire after they had completed rating all of the videos (Table 4). Raters were asked to give feedback on the tool, including whether they found it to be unclear, difficult to use, or any other inputs. The answers are summarized completely in Table 4.

Consequences

We examined the possibility of using NorMS-NTS after minimal rater training. For a high-stake summative assessment, an ICC of above 0.70 is suggested [31]. For a formative assessment, a minimum ICC is not clearly specified. An ICC above 0.60, however, is proposed [31]. The proposed ICC levels are based on the average ICC. The average ICC levels are always higher than the single-measure ICCs [32]. We could not find any proposed levels for single ICC measures for formative assessment.

Results

The average overall global scores for the three raters across the 20 videos was 4.7 (SD = 1.1), 4.3 (SD = 1.4) and 4.0 (SD = 2.0).

Table 4 Raters questionnaire

Background:
Age:
Specialization:
Academic competency highest degree/position?
Clinical experience (number of years in clinical practice):
Do you have any prior experience with nontechnical skills (NTS) or tools for NTS assessment?
Usability of the tool:
How was the tool to use?
How easy was it to assess the students' skills in elements and categories?
Were there elements of nontechnical skills that the tool did not capture?
Were there elements that you felt were redundant, i.e., should not have been included in the tool?
Were there elements that were difficult to assess?
Were the written instructions helpful?
Did you find that it became easier or more difficult to use the tool after gaining more experience with its use?
How long did you spend on average rating the videos?
Is this a tool you could use for training or teaching?
Other feedback?

Internal structure

Interrater reliability

An ICC below 0.40 is considered as a poor correlation, between 0.40 and 0.59 is considered a fair correlation, between 0.60 and 0.74 is considered an good correlation and between 0.75 and 1.00 as excellent correlation [33]. The ICC agreement for the sum score of the overall global score for all raters was fair: ICC (3,1) = 0.53 [33]. This was supported by Kendall's W = 0.73 (Table 5). Two of the raters had a higher level of experience, and once an agreement analysis for those two only was applied, the level of agreement was higher. ICC (3,1) = 0.53 was still fair [33]; however, Kendall's W = 0.80 was good. The individually calculated ICC (3,1) and Kendall's W are both lower (0.25–0.55 and 0.51–0.75, respectively).

Internal consistency analysis

For both the Spearman correlation coefficient and Cronbach's alpha, a correlation coefficient of near 1.0 represents high internal consistency. Most of the Spearman correlations were above 0.80 (Table 6). The correlation coefficients for the pooled raters were in the range of 0.77–0.91. Almost all correlation coefficients were significant at the $p = 0.0001$ level. Cronbach's alpha for the elements, categories and global scores were all mostly above 0.90, which is in the excellent range and thus confirms a high level of scoring consistency among the raters.

Observability

Observability was calculated as the percentage of elements and categories that were not scored with n/a. Two of the marking forms had completed scoring of all elements scored but not all categories. This was considered an error, as all elements were observed. Those two forms were not included in the statistics. The observability was deemed acceptable (95%-100%) (Table 7).

Response process

The raters' responses are summarized in Table 8. All the raters found the tool easy to use, none of the elements were identified as redundant, and the written instructions were helpful. Raters also found the tool easier to use once they gained practice in using it. Raters with NTS experience had a shorter time of use per video than the novel rater. All the raters stated that they could use the tool for training or teaching.

Some of the videos were reported to be slightly too brief to properly assess all elements for scoring. One of the raters suggested that the ratings should have been more standardized, that team members should be more uniform and that facilitators should take a similar approach. It was also mentioned that communication depended on whether the patient was a manikin or a simulated patient.

Table 5 Inter-rater agreement statistics. ICC and Kendall's W

Score	All raters		Rater 2 and 3	
	ICC(3,1)	Kendall's W	ICC(3,1)	Kendall's W
Communication	0.49	0.69	0.37	0.71
Team communication	0.43	0.63	0.48	0.77
Establish mutual communication	0.55	0.75	0.45	0.80
Patient communication	0.54	0.68	0.45	0.74
Situational awareness	0.50	0.69	0.43	0.73
Situational assessment	0.27	0.51	0.07	0.56
Understanding of team members' roles	0.39	0.63	0.13	0.58
Attentiveness	0.44	0.68	0.37	0.76
Teamwork	0.40	0.62	0.20	0.63
Professional modesty	0.25	0.51	0.02	0.55
Flexibility	0.41	0.67	0.40	0.76
Efficient use of team members	0.40	0.62	0.25	0.64
Decision making	0.44	0.68	0.49	0.79
Uncertainty management	0.36	0.57	0.46	0.75
Decision analysis	0.43	0.61	0.58	0.81
Leadership	0.49	0.72	0.48	0.82
Prioritization	0.33	0.56	0.37	0.71
Overall Global Score	0.53	0.73	0.55	0.80
Sum of communication elements	0.58	0.76	0.51	0.81
Sum of situational awareness elements	0.41	0.67	0.21	0.67
Sum of teamwork elements	0.42	0.68	0.28	0.71
Sum of decision-making elements	0.46	0.66	0.55	0.82
Sum of all elements	0.50	0.72	0.45	0.82
Sum of categories	0.52	0.73	0.45	0.80

One rater suggested that crew resource management (CRM) elements, such as fixation errors and reevaluations, could be given a greater focus in the tool.

Consequences

The calculations show that the use of NorMS-NTS by raters new to the tool reaches an ICC of 0.53. That value is 0.08 below the suggested ICC level for formative assessment of above 0.60 [34].

Discussion

The NorMS-NTS tool was developed for the assessment of Norwegian medical students' nontechnical skills. Our aim has been to create an easy-to-use tool that suits busy doctors as near-peer educators in both clinical teaching settings and during simulation-based training. Ideally, this tool should be easy to find online, and raters should be able to use the tool after only a short introduction. The interpretation of the validation results described in this study was based on these principles.

The raters found the tool usable. They found all the categories and elements relevant. The raters considered the written instructions helpful. We will improve them further, especially for the categories and elements with the lowest ICC. All raters could use the tool for training or teaching. The least experienced rater used 45 min to rate videos, which is not feasible in clinical practice, bu the experienced raters used only a few minutes more than the duration of the scenario. Therefore, raters will probably be more efficient as they become accustomed to the tool. The raters also described that in their feedback. The internal structure of the tool was excellent. The observability was also found to be excellent. These findings support the tools' structure and content. The usability of the tool was found to be satisfactory.

The usability for the raters after only a short introduction is an important part of the 'Consequences'. On the other hand, the consequences for the students are also important to investigate further. Such studies should explore the students' views. Are they assessed fairly? Do they get ideas for improvement? Does the assessment motivate or encourage them? It is also important to explore the system consequences. Is it possible to integrate such a tool in education? Do teachers and learners use the tool to clarify learning potential, or a test to pass or fail. Do we have the tools to help those who struggle?

Table 6 Consistency in scoring by Spearman correlation coefficient for category vs. elements or global score vs. categories

Score	Spearman correlation coefficient for category vs. elements or global score vs. categories				Cronbach's alpha (standardized variables)			
	Rater 1	Rater 2	Rater 3	Raters pooled	Rater 1	Rater 2	Rater 3	Raters pooled
Communication	-	-	-	-	0.94	0.74	0.90	0.92
Team communication	0.88	0.85	0.89	0.91	0.96	0.75	0.92	0.93
Establish mutual communication	0.89	0.68	0.85	0.86	0.96	0.84	0.92	0.94
Patient communication	0.90	0.43	0.77	0.82	0.95	0.87	0.96	0.95
Situational awareness	-	-	-	-	0.95	0.52	0.98	0.94
Situational assessment	0.96	0.47	0.91	0.88	0.95	0.84	0.98	0.96
Understanding of team members' roles	0.88	0.82	0.91	0.87	0.97	0.66	0.98	0.96
Attentiveness	0.93	0.58	0.96	0.89	0.96	0.74	0.97	0.95
Teamwork	-	-	-	-	0.92	0.77	0.97	0.94
Professional modesty	0.85	0.71	0.97	0.88	0.94	0.80	0.98	0.95
Flexibility	0.92	0.52	0.97	0.88	0.93	0.88	0.98	0.95
Efficient use of team members	0.87	0.73	0.95	0.89	0.96	0.79	0.99	0.96
Decision making	-	-	-	-	0.94	0.88	0.97	0.95
Uncertainty management	0.88	0.85	0.89	0.87	0.95	0.88	0.97	0.95
Decision analysis	0.88	0.76	0.82	0.80	0.96	0.91	0.98	0.96
Leadership	0.92	0.57	0.92	0.86	0.95	0.94	0.97	0.96
Prioritization	0.87	0.90	0.95	0.90	0.95	0.89	0.97	0.95
Overall global score	-	-	-	-	0.94	0.82	0.98	0.95
Communication	0.88	0.82	0.90	0.81	0.96	0.84	0.98	0.95
Situational awareness	0.93	0.74	0.89	0.77	0.95	0.85	0.98	0.95
Teamwork	0.86	0.55	0.94	0.79	0.97	0.90	0.98	0.95
Decision making	0.94	0.69	0.93	0.86	0.95	0.86	0.98	0.95

Table 7 Observability

Elements	Observability
Team communication	100%
Establish mutual understanding	100%
Patient communication	100%
Situational assessment	100%
Understanding of team members' roles	100%
Attentiveness	100%
Professional modesty	100%
Flexibility	100%
Efficient use of team members	100%
Decision analysis	97%
Uncertainty management	95%
Leadership	98%
Prioritization	100%
Categories	
Communication	100%
Situational awareness	100%
Teamwork	100%
Decision making	99%

This is all out of scope for this paper but should be studied further.

The individual interrater reliability after a short introduction and training was found to be fair. We found a single measure ICC of 0.53 for the global overall score. That ICC is 0.08 below the suggested ICC level for formative assessment (above 0.60) [34]. However, we know that the suggestion is based on the average ICC, which is always higher than a single-measure ICC [35]. Comparing to other NTS tools, ICC is challenging, as the ICC calculations are not specified [36]. In studies where single-measure ICC is calculated with raters novice to the tool the findings are quite similar to ours. The NOTSS single measure ICCs on the category scores varied from 0.29 to 0.66 [37]. The Medi-StuNTS reached a single-measure ICC of 0.37 [36]. Other studies where ICC is not specified as single-measures or average the ICC are still in the same range as NorMS-NTS [38]. A study comparing ANTS and Ottawa GRS found ICCs of 0.39 and 0.42 for overall scores [39]. As there are no suggested levels for single-measure ICCs for formative assessment for novice raters [36], we consider the calculated levels to be sufficient for conducting a formative assessment of medical student NTS, as they are in the same range as for

Table 8 Rater feedback

Usability	Rater 1	Rater 2	Rater 3
How was the tool to use?	*Quite intuitive*	*I think the tool seemed adequate in relation to the purpose*	*I think the form actually works quite well*
How was it to assess the students' skills in elements and categories?	*A bit detailed and challenging to get all the elements, but absolutely possible.*	*I think the tool will be able to work well. I think the evaluation of the form would have done better with a more uniform background among team members in the scenario, standardization in the use of the tool and a similar approach among facilitators*	*There are good points that are assessed under each main category, but I am curious to see how much our assessments match or differ from each other when comparing the results between us users*
Were there elements of non-technical skills that the tool did not capture?	*Not really, but the videos could sometimes be a little short so that you couldn't include, for example, decisions at the end. And probably better to assess candidates when you are physically present*	*CRM elements such as fixation errors and reevaluation could perhaps be given their own focus?*	*No*
Were there elements that you felt were redundant, i.e. should not have been included in the tool?	*Not the way I see it, all elements are important*	*No*	*No*
Were there elements that were difficult to assess?	*Modesty must be appropriate - too much, then maybe it will also be difficult with the cooperation*	*It becomes difficult to assess communication with the patient depending on whether a doll or living marker is used*	*No, but perhaps a little overlap between the categories*
Were the written instructions helpful?	*Yes*	*Yes - as an assurance that you had the correct understanding of the form*	*Yes*
Did you find that it became easier or more difficult to use the tool as you used it more times?	*Yes, easier once you get used to it*	*Yes*	*Easier*
How long did you spend on average rating the videos?	*I went through all the videos and reassessed twice. to see if I rated roughly the same. I would estimate about 45 minutes per video, so you have to have breaks.*	*That time the video lasted + 2-3 minutes*	*I have mostly had one review per candidate. The way I have done this was to fast forward to the arrival of the doctor and follow the scenario from there. 20-30 minutes?*
Is this a tool you could use for training or teaching?	*Yes*	*Yes*	*Yes, I think it can be useful, but with a clear order of what to use it for.*
Other feedback?			*There is a wide gap in both candidates and facilitators in terms of competence, experience and skills. I think that is an important limitation of the study. The candidates who get to appear as a marker necessarily have a slightly more difficult job with communication than those who have a living marker. It was a long time from when I completed this work*

other validated NTS tools. The average ICC (3.1) would be more appropriate to use for validation for summative assessment and should be applied in later validations of the tool.

There are several ways to increase interrater reliability, i.e., rater training, modification of the assessment tool, stricter scenario design, etc. Previous studies have shown that the level of interrater agreement increases when raters gain more experience with an assessment tool [40]. As the NorMS-NTS is usable with minimal training, it is also possible for busy doctors to gain experience with the tool, hence increasing its interrater reliability. We will also continue to refine the NorMS-NTS training introduction and training manual in the areas that were identified as poor.

Limitations

As collecting validity evidence of NTS assessment tools is a continuous process of collecting evidence of validity, this article only describes part of the validation necessary to meet all accepted sources of evidence in the Messick framework. We have tried to clearly specify the context and intended use we have assessed usability and preliminary validation of NorMS-NTS for in this article. We did not seek validity evidence of the use of the tool for summative assessment with minimal rater training now. Further collection of validity evidence as described in the Messick framework is planned, including for summative assessment using average ICC. To fully validate the tool for formative assessment, it is necessary to further study the consequences of the tool. That is, we explore the impact on the students and see if the formative assessments obtained by the tool are correct and beneficial.

The raters had some input about the validation process itself. We deliberately chose to not have standardized scenarios, teams, and facilitators. We wanted a tool that works in everyday life, with different facilitators, team members and situations. All raters rated the same scenarios in the study, so they had the same variety. We would probably have achieved a higher level of interrater reliability with a greater degree of standardization of the scenarios and ratings, but the findings may not have been transferable to practical use. Some suggest that all validation of assessment tools should include true measurers of validity and reliability, and we have worked to achieve this in our study [41].

As this preliminary validation process was created to validate the tool for formative assessment for busy doctors as near-peer educators in clinical practice, we chose single-measure ICCs. Because of that, we only had three raters. When validating the tool for summative assessment, more raters will be included.

The tool was developed in Norway. When using it in different contexts, be it different places within Norway or in different countries, pilot studies should be conducted, collecting context-specific validity evidence again. Using such a tool and interpreting its results is a complex socio-technical endeavor with possible consequences for healthcare professionals and the people who they treat. Therefore, it seems appropriate to double check.

Conclusions

We collected preliminary evidence of validity for the NorMS-NTS tool. Raters found the tool usable. When the NorMS-NTS was used by raters new to the tool we found that the interrater reliability, internal consistency, and observability were sufficient for formative assessment. It is necessary to further examine the consequences of the tool to fully validate the tool for formative assessment.

Further

The process of validation for the NorMS-NTS began with this study. A summative assessment study calculating the average ICC is planned for the future. Further validation should focus on the final two sources of evidence in the Messick framework: relationship with other variables and consequences. We note that it is also important to validate the tool for different settings.

Abbreviations

NorMS-NTS	Norwegian Medical Student's Non-technical Skills
NTS	Non-technical skills
TSD	Services for sensitive data

Acknowledgements

N/a.

Authors' contributions

KP made all the video recordings, and she performed the rater training. KP transferred the data from the rating forms to Excel. HF performed the statistical analysis based on the Excel spreadsheet. TW, PD and DM were all major contributors in planning and writing the manuscript. All the authors have read and approved the final manuscript.

Authors' information

David M Musson currently is an Associate Professor at McMaster university and the Northern Ontario School of Medicine. He does research in space medicine, decision support, human factors, expert systems, applied social psychology, and medical training.
Peter Dieckmann is Senior Scientist at CAMES, Professor of Healthcare Education and Patient Safety, University of Stavanger, Norway. and External Lecturer at the University of Copenhagen.
Peter Dieckmann's research on simulation addresses the theoretical foundations and practice of using simulation to create, recognize and use learning opportunities for simulation learners and faculty members in different contexts. Peters research with simulation focuses on exploring simulation's possibilities to optimize work systems. Beyond simulation, Peter investigates issues of the practice of patient safety work, psychological safety, medication safety, and visual research methods.
Torben Wisborg is a professor of anesthesia and consultant anesthetist. He is also research director for the Norwegian National Advisory Unit on Trauma. He did his PhD on trauma team training based on ten years of experience training trauma teams in Norway and northern Iraq. Main research interests are trauma systems, trauma teams and trauma team training, including simulation.

Funding

Open access funding provided by UiT The Arctic University of Norway (incl University Hospital of North Norway) This work was supported by the Northern Norway Regional Health Authority (Ref: HNF1368-17).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Norwegian law exempts educational studies from ethical approval because they do not involve patients. However, the Regional Committee of North Norway for Medical and Health research gave feedback on the protocol and approved this assumption (Ref: 2016/1539/REK nord). The consent form for the participants received approval by the Norwegian Center of Research Data (Ref: 57474/2017). Informed consent from all participants was obtained after oral and written information was delivered on the purpose and objectives of the study. All videos were saved at the TSD (Tjeneste for Sensitive Data) facilities owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT department (USIT). (tsd-drift@usit.uio.no. TSD provides a platform for public research institutions in Norway. This service provides a secure project area where researchers can collect, store, and analyze sensitive data.

Consent for publication

N/a.

Competing interests

The authors declare no competing interests.

Author details

[1]Interprofessional Rural Research Team, Faculty of Health Sciences, Department of Clinical Medicine, University of Tromsø – the Arctic University of Norway, Hammerfest, Norway. [2]Hammerfest Hospital, Finnmark Health Trust, Hammerfest, Norway. [3]Copenhagen Academy for Medical Education and Simulation (CAMES), Center for Human Resources and Education, Capital Region of Denmark, Copenhagen, Denmark. [4]Faculty of Health Sciences, Department of Quality and Health Technology, University of Stavanger, Stavanger, Norway. [5]Department of Public Health, Copenhagen University, Copenhagen, Denmark. [6]Capturo AS, Skjetten, Norway. [7]Faculty of Health Sciences, Department of Anesthesia, McMaster University, Hamilton, ON, Canada.

Received: 2 May 2023 Accepted: 2 November 2023
Published online: 15 November 2023

References

1. Flin RH, O'Connor P, Crichton M. Safety at the sharp end: a guide to non-technical skills. England, U.K: Ashgate; 2008.
2. Jepsen RMHG, Østergaard D, Dieckmann P. Development of instruments for assessment of individuals' and teams' non-technical skills in healthcare: a critical review. Cogn Technol Work. 2015;17(1):63–77.
3. Hagemann V, Herbstreit F, Kehren C, Chittamadathil J, Wolfertz S, Dirkmann D, Kluge A, Peters J. Does teaching non-technical skills to medical students improve those skills and simulated patient outcome? Int J Med Educ. 2017;8:101–13.
4. Somasundram K, Spence H, Colquhoun AJ, McIlhenny C, Biyani CS, Jain S. Simulation in urology to train non-technical skills in ward rounds. BJU Int. 2018;122(4):705–12.
5. Yeung J, Husselbee N, Davies R, Melody T, Lockey A, Gwinutt C, Bullock I, Gao F, Perkins G. Effect of non-technical skills teaching on performance – an evaluation of additional non-technical skills teaching in resuscitation council ALS course. Resuscitation. 2015;96:24.
6. Nicolaides M, Theodorou E, Hanrahan JG, Theodoulou I, Emin EI, Papalois A, Sideris M. Advancing medical students' non-technical skills in a group-based setting. Journal of investigative surgery : the official journal of the Academy of Surgical Research. 2021;34(1):39–43.
7. Donaldson L, Ricciardi W, Sheridan S, Tartaglia R. Textbook of patient safety and clinical risk management. Cham, Switzerland: Springer International Publishing; 2020.

8. Fletcher GC, McGeorge P, Flin RH, Glavin RJ, Maran NJ. The role of non-technical skills in anaesthesia: a review of current literature. Br J Anaesth. 2002;88(3):418–29.
9. Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Rating non-technical skills: developing a behavioural marker system for use in anaesthesia. Cogn Technol Work. 2004;6(3):165–71.
10. Byrne AJ, Greaves JD. Assessment instruments used during anaesthetic simulation: review of published studies. Br J Anaesth. 2001;86(3):445–50.
11. Glavin RJ, Maran NJ. Development and use of scoring systems for assessment of clinical competence. Br J Anaesth. 2002;88(3):329–30.
12. Jepsen RM, Dieckmann P, Spanager L, Lyk-Jensen HT, Konge L, Ringsted C, Østergaard D. Evaluating structured assessment of anaesthesiologists' non-technical skills. Acta Anaesthesiol Scand. 2016;60(6):756–66.
13. Schuwirth L, Colliver J, Gruppen L, Kreiter C, Mennin S, Onishi H, Pangaro L, Ringsted C, Swanson D, Van Der Vleuten C, et al. Research in assessment: consensus statement and recommendations from the Ottawa 2010 conference. Med Teach. 2011;33(3):224–33.
14. Moll-Khosrawi P, Kamphausen A, Hampe W, Schulte-Uentrop L, Zimmermann S, Kubitz JC. Anaesthesiology students' Non-Technical skills: development and evaluation of a behavioural marker system for students (AS-NTS). BMC Med Educ. 2019;19(1):205.
15. Lyk-Jensen HT, Jepsen RM, Spanager L, Dieckmann P, Østergaard D. Assessing Nurse Anaesthetists' Non-Technical Skills in the operating room. Acta Anaesthesiol Scand. 2014;58(7):794–801.
16. Yule S, Flin R, Paterson-Brown S, Maran N, Rowley D. Development of a rating system for surgeons' non-technical skills. Med Educ. 2006;40(11):1098–104.
17. Rutherford JS, Flin R, Irwin A, McFadyen AK. Evaluation of the prototype anaesthetic non-technical skills for anaesthetic practitioners (ANTS-AP) system: a behavioural rating system to assess the non-technical skills used by staff assisting the anaesthetist. Anaesthesia. 2015;70(8):907–14.
18. Mitchell L. Scrub practitioners' list of intra-operative non-technical skills – SPLINTS. 1st ed. Boca Raton, F.L: CRC Press; 2009.
19. Hamilton AL, Kerins J, MacCrossan MA, Tallentire VR. Medical Students' Non-Technical Skills (Medi-StuNTS): preliminary work developing a behavioural marker system for the non-technical skills of medical students in acute care. BMJ Simulation and Technology Enhanced Learning. 2019;5(3):130–9.
20. Gourbault LJ, Hopley EL, Finch F, Shiels S, Higham H. Non-technical Skills for Medical Students: Validating the Tools of the Trade. Cureus. 2022;14(5):e24776.
21. Jepsen RM, Spanager L, Lyk-Jensen HT, Dieckmann P, Østergaard D. Customisation of an instrument to assess anaesthesiologists' non-technical skills. Int J Med Educ. 2015;6:17–25.
22. Spanager L, Lyk-Jensen HT, Dieckmann P, Wettergren A, Rosenberg J, Ostergaard D. Customization of a tool to assess danish surgeons´ non-technical skills in the operating room. Dan Med J. 2012;59(11):A4526.
23. Prydz K, Dieckmann P, Musson D, Wisborg T. The development of a tool to assess medical students' non-technical skills - the Norwegian medical students' non-technical skills (NorMS-NTS). Med Teach. 2023;45(5):516–23.
24. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. Advances in simulation (London, England). 2016;1:31.
25. Williams D HR, Boehm-Davis D. : Proceedings of the 9th International Symposium on Aviation Psychology; Training for inter-rater reliability: baselines and benchmarks. ; 1997.
26. American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, Psychological Testing: Standards for educational and psychological testing. Washington , D.C: American Educational Research Association; 2014.
27. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. Adv Simul. 2016;1(1):31.
28. Yule S, Flin R, Maran N, Rowley D, Youngson G, Paterson-Brown S. Surgeons' non-technical skills in the operating room: reliability testing of the NOTSS behavior rating system. World J Surg. 2008;32(4):548–56.
29. Hamilton AL, Kerins J, MacCrossan MA, Tallentire VR. Medical students' non-technical skills (Medi-StuNTS): preliminary work developing a behavioural marker system for the non-technical skills of medical students in acute care. BMJ simulation & technology enhanced learning. 2019;5(3):130–9.
30. Abell N, Springer DW, Kamata A. Developing and validating rapid assessment instruments. Oxford, U.K: Oxford University Press; 2009.
31. Hull L, Arora S, Symons NR, Jalil R, Darzi A, Vincent C, Sevdalis N. Training faculty in nontechnical skill assessment: national guidelines on program requirements. Ann Surg. 2013;258(2):370–5.
32. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420–8.
33. Cicchetti DV. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. Psychol Assess. 1994;6(4):284–90.
34. Hull L, Arora S, Symons NR, et al. Training faculty in nontechnical skill assessment: national guidelines on program requirements. Ann Surg. 2013;258(2):370–5.
35. LeBreton JM, Senter JL. Answers to 20 questions about interrater reliability and interrater agreement. Organ Res Methods. 2008;11(4):815–52.
36. Clarke B, Smith SE, Phillips EC, Hamilton A, Kerins J, Tallentire VR. Reliability of assessment of medical students' non-technical skills using a behavioural marker system: does clinical experience matter? BMJ Simul Technol Enhanc Learn. 2021;7(5):285–92.
37. Yule S, Flin R, Maran N, Rowley D, Youngson G, Paterson-Brown S. Surgeons' Non-technical Skills in the Operating Room: Reliability Testing of the NOTSS Behavior Rating System. World J Surg. 2008;32(4):548–56.
38. Graham J, Hocking G, Giles E. Anaesthesia non-technical skills: Can anaesthetists be trained to reliably use this behavioural marker system in 1 day? Br J Anaesth. 2010;104(4):440–5.
39. Nunnink L, Foot C, Venkatesh B, Corke C, Saxena M, Lucey M, Jones M. High-stakes assessment of the non-technical skills of critical care trainees using simulation: feasibility, acceptability and reliability. Crit Care Resusc. 2014;16(1):6–12.
40. Walsh DP, Chen MJ, Buhl LK, Neves SE, Mitchell JD. Assessing interrater reliability of a faculty-provided feedback rating instrument. J Med Educ Curric Dev. 2022;9:23821205221093204.
41. Van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J. Objective assessment of technical surgical skills. Br J Surg. 2010;97(7):972–87.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Paper 3

Non-technical skills of Norwegian medical students at different training sites: A comparative, observational cohort study

Katrine Prydz[a,b,*], Peter Dieckmann[c,d,e], Hans Fagertun[f], David Musson[g], Torben Wisborg[a,b]

[a]Interprofessional Rural Research Team, Faculty of Health Sciences, Department of Clinical Medicine, University of Tromsø – the Arctic University of Norway, Hammerfest, Norway

[b]Hammerfest Hospital, Finnmark Health Trust, Hammerfest, Norway

[c]Copenhagen Academy for Medical Education and Simulation (CAMES), Center for Human Resources and Education, Capital Region of Denmark, Copenhagen, Denmark and University of Stavanger, Norway

[d]Faculty of Health Sciences, Department of Quality and Health Technology, University of Stavanger, Norway

[e]Department of Public Health, Copenhagen University, Denmark

[f]Capturo AS

[g]Faculty of Health Sciences, Department of Anesthesia, McMaster University, Ontario, Canada

[*]Corresponding author

Katrine Prydz

Finnmark Hospital Trust

Faculty of Health Sciences, Department of Clinical Medicine, UiT – the Arctic University of Norway, Hvalrossveien 12, 9602 Hammerfest, Norway

Email: katrine.prydz@uit.no

Abstract

Purpose:

Mastering non-technical skills (NTS) is a fundamental part of the training of new physicians to perform effectively and safely in the medical practice environment. Ideally, they learn these skills during medical school. Decentralized medical education is being implemented increasingly worldwide. This study aimed to compare the NTS performance of medical students in their last year of education at three different training sites of the same university. Two of the three training sites studied, Bodø (a regional hospital) and Finnmark (a rural local hospital), implemented decentralized medical education. The third training site was the main campus in Tromsø, located at an urban university hospital.

Methods:

This blinded cohort study included students from the three training sites who participated in identical multi-professional simulations over a six-year period. Eight raters evaluated the video recordings of eight students from each training site using the Norwegian Medical Students Non-Technical Skills (NorMS-NTS) tool. The NorMS-NTS tool, which comprises four categories and 13 elements, assesses the NTS of Norwegian medical students and assigns an overall global score. Pairwise significant differences in the NTS performance levels between the training sites studied were assessed using Tukey's test.

Results:

The overall NTS performance levels of the medical students from Finnmark (mean 4.5) were significantly higher than those of the students from Tromsø (mean 3.8) and Bodø (mean 3.5). Similarly, the NTS performance levels at category-level of the students in Finnmark were significantly higher than those of the students from Bodø and Tromsø. Except for one

category, no significant differences were observed between the students from Bodø and Tromsø in terms of the overall or category-level NTS performance.

Conclusion:

The NTS performance levels of the medical students from Finnmark, which implements rural, decentralized medical education, were significantly higher than those of the students from Tromsø and Bodø.

Introduction

The first medical school in Northern Norway, established at UiT – the Arctic University of Norway (UiT) 50 years ago,(1) was the first rural-oriented medical education model in Europe to recruit physicians to the underserved population of northern Norway with the intent to improve the health care standards in the region.(1) The university implemented one year of training in rural general practice and local hospitals outside the campus.(1) Moreover, the university also prioritized applicants from Northern Norway(1) owing to an expected Salmon Effect, which hypothesizes that physicians, similar to salmons, return to the region they grew up.(2) A previous study concluded that this education model is sustainable and can facilitate the recruitment of physicians to the northern regions.(1)

In 2009 UiT developed *the Bodø model*,(3) a decentralized model wherein 24 medical students from UiT completed the sixth and last year of undergraduate medical education in Bodø, in addition to the placement in the fifth year of study.(4) The Bodø model aimed to address the limitations of clinical training capacity available in Tromsø.(4) Bodø is home to the second largest hospital in Northern Norway, and these students are located at that hospital.(4) This model was developed on the principle that the students followed the same schedule as that at Tromsø.(3) Academic training schedules at the two centers are largely similar, with only minor variations in learning activities. In 2017 UiT developed *the Finnmark model for medical student training*, a decentralized model wherein students complete the fifth and sixth year of medical school in the rural county of Finnmark rather than the main training site in Tromsø.

To ensure consistency in medical education, the quality of teaching must be assessed in decentralized education. Students in Bodø, Tromsø, and Finnmark undertake a common final exam. Examination results can be used to assess outcomes of students' learnings at different training sites. Reports from UiT for the period from 2018–2023 revealed that 87.3% of the

490 students in Tromsø successfully passed the final exam. Moreover, the passing percentage of the students who received decentralized education was higher (91.2% of the 147 students and 96% of the 50 students in Bodø and Finnmark, respectively).

The Finnmark model places an extensive focus on the acquisition of non-technical skills (NTS) via the continuous use of simulations so that the students may achieve high levels of NTS. These NTS includes situational awareness, decision-making, communication, teamwork and leadership.(5) Previous studies have demonstrated the role of NTS of the health care professionals in patient safety.(6) Insufficient NTS have been identified as a contributing factor in 70% of adverse events occurring in hospital settings.(7) NTS include interpersonal skills and complement the necessary technical skills required for clinical practice.(8) In contrast, technical skills are the profession-specific competency possessed by health professionals and students.(5, 9)

Researchers have debated the use of the term NTS.(10) Nevertheless, NTS remains the most commonly used description. NTS can be acquired via training,(11) and higher levels of NTS have been shown to improve patient safety.(5, 12) Therefore, health professionals and students should undergo NTS training.(13)

Evaluating the outcomes of training is an essential element of high-quality training. As, providing feedback to students and health professionals on their NTS levels will aid in increasing the focus on gaining the right skills, it is necessary to develop tools to assess the NTS of health professionals and students. Tools have been used to assess NTS in the field of aviation for decades(14), and have been developed to assess the NTS of health professionals since the beginning of 2000.(8, 15) Previously, we have developed NorMS-NTS, a tool that assesses the NTS of Norwegian medical students, in 2022 .(16)

This study aimed to assess the possible differences in NTS performance levels among synchronous groups of medical students from different training sites by comparing the NTS performance of the decentralized students in rural Finnmark and the city of Bodø with those of the students at the main training site in Tromsø.

Methods

Overview

This was an observational cohort study.(17) The three cohorts were comprised of eight medical students from each of the training sites studied. The experimental variable was a different training site. Eight raters who were blinded to the training site assessed the NTS performance of the 24 medical students using the NorMS-NTS tool (Table 1). We compared the results of the statistical analysis thereafter.

Table 1. The NorMS-NTS tool

Category*	Category score**	Element*	Element score**	Feedback
Communication		Team communication		
		Establish mutual understanding		
		Patient communication		
Situation awareness		Situational assessment		
		Understanding of team members' roles		
		Attentiveness		
Teamwork		Professional modesty		
		Flexibility		
		Efficient use of team members		
Decision making		Uncertainty management		
		Decision analysis		
		Leadership		
		Prioritization		

General comments:

____

____

____

____

*N/A - Not applicable. 1, much below average; 2, below average; 3, acceptable; 4, above average; 5, much above average.

** Within team unless other specified.

Overall global rating (Mark with a ring)**:**

Very poor 1 – 2 – 3 – 4 – 5 – 6 – 7 Excellent

Setting

The students at all three UiT training sites studied participate in InterSim, a simulation-based multi-professional training program that encompasses different acute care situations with standardized scenarios, during the last term of the sixth year of their undergraduate program. The medical students were paired with nursing students forming teams, which also included radiographers and bioengineering students in some cases.

We video recorded the sessions of the medical students from Finnmark, Bodø, and Tromsø participating in two different scenarios, with each scenario lasting 12–20 minutes. The first scenario involved a patient with sepsis, whereas the second scenario involved a patient with postoperative dyspnea. One or two trained physicians and nurses facilitated each simulation and debriefing. Although most scenarios employed a simulated patient, some scenarios were performed using a simulation manikin owing to the COVID-19 restrictions. The students performed all measurements, examinations, and tests on the patients and gathered details regarding pain and emotions. The facilitators provided the answers subsequently. The students informed the facilitators regarding the procedures they aimed to conduct if the equipment was missing, and the facilitator provided the results.

Video recordings

All participants were sixth-year medical students from Finnmark, Bodø, or Tromsø. We provided a thorough explanation regarding the objectives and aims of the study to the participants. Participation was voluntary and had no consequences for their education. We randomly selected eight video recordings with sufficient sound and image quality from the video recordings of over 100 teams acquired between 2018 and 2023. We designated identification numbers to the videos and randomized their order.

Raters

We contacted two national simulation networks (InterRegSim, a national collaboration for simulation-based learning in the specialist health service in Norway(18) and the Better & Systematic Team training [BEST] network, an international multi-professional team training program originating in Finnmark(19)) via email to recruit raters. We offered gift cards of NOK 3000 to the 11 raters recruited. We recruited more raters than necessary to compensate for dropouts. Eight raters, comprising four men and four women aged 46–69 years (mean: 55 years), completed the task within a specified timeframe.

The raters had 22–44 (mean: 27.5) years of clinical experience; two raters did not answer this question. Six raters were registered nurses, whereas two were medical doctors. Seven raters reported prior experience with the NTS and/or NTS tools; the last rater did not answer this question.

All raters familiarized themselves with the NorMS-NTS tool and received a presentation of the tool from a researcher (KP) via Microsoft Teams. All raters received secure online access to the 24 videos. We assigned a two-digit study identification code to each video and a number to each rater. The raters who were blinded to the training sites studied rated all videos using the NorMS-NTS tool and returned the ratings to a researcher (KP) via e-mail. The raters were only aware of the identification numbers of the students.

Sample size

The NorMS-NTS is a new tool, and no previous study has assessed the NTS of Norwegian medical students. Consequently, we could not obtain any estimates of the prevalence or standard deviation or calculate the sample size.(20) The practical implications of the selected sample size were also important. Three raters rated 20 videos of the same length as those used in the present study in a previous study(21) and exhibited a nearly identical threshold for the number of videos each rater could rate. The assessment process is time-consuming and requires a focused rater. Therefore, we selected 24 videos and increased the number of raters to eight to increase the likelihood of accurately measuring the NTS performance levels.(22)

The NorMS-NTS tool

The NorMS-NTS tool used in this study was developed to assess the NTS of Norwegian medical students via observation in 2022 (Table 1).(16) This tool consists of four categories comprising 13 elements rated on a 5-point Likert scale. The global overall score is rated on a 7-point Likert scale.

Statistical Analysis

We analyzed all data extracted from the NorMS-NTS forms of the raters using Statistical Analysis System (SAS 9.4). We compared the mean element-level, category-level, and overall NTS performance levels at different training sites subsequently. The null hypothesis was that no significant differences would be observed among the NTS performance of the students in the three cohorts. Tukey's test (23) which is a test that adjust for type I-error was used to calculate significant difference.(24)

Ethics

Educational studies that do not involve patients are exempt from ethical approval in Norway. The Regional Committee of North Norway for Medical and Health Research waived the requirement for a formal review of this study protocol (Ref: 2016/1539/REK nord) in 2016.

The Norwegian Centre for Research Data (NSD) was entrusted as the Data Protection Official for Research of the Finnmark Hospital Trust and approved the project in 2017 (NSD Ref: 57474/2017). The NSD ensures legal access to necessary personal data for research and provides data-protection services to all Norwegian universities.

We provided written and oral explanations about the purpose and objectives of the study to the students and obtained informed consent from all participants.

Services for sensitive data (TSD) at the University of Oslo provide a platform for public research institutions in Norway. Researchers can collect, store, and analyze sensitive data in a secure project area. We stored all videos used in this study on a TSD platform. We also conducted the rating of the students at TSD facilities.

Results

Overall NTS performance

The NTS performance levels of the medical students from Finnmark (mean 4.52 (0.25)) were significantly higher than those of the students from Bodø (mean 3.53 (0.25)) and Tromsø (mean 3.83 (0.25)) on a scale ranging from 1 to 7 (Table 2 and Figure 1). However, we observed no significant difference between the NTS performance levels of the medical students from Tromsø and Bodø.

Table 2. Observed differences in NTS performance between the three training sites. Significant when p<0.05.

Category/element	Finnmark	Tromsø	Bodø	Finnmark-Bodø		Finnmark-Tromsø		Tromsø-Bodø	
	Mean (SD)	Mean (SD)	Mean (SD)	Mean diff. (95% CI)	P-value	Mean diff. (95% CI)	P-value	Mean diff. (95% CI)	P-value
Overall score	4.52 (0.25)	3.83 (0.25)	3.53 (0.25)	0.99 (0.23 to 1.75)	<.0001	0.69 (-0.14 to 1.53)	0.0023	0.30 (-0.46 to 1.06)	0.31
Category									
Communication	3.59 (0.15)	3.01 (0.15)	2.96 (0.15)	0.63 (0.18 to 1.09)	<.0001	0.58 (0.09 to 1.07)	0.0001	0.05 (-0.39 to 0.50)	0.92
Situational assessment	3.58 (0.15)	3.26 (0.15)	3.04 (0.15)	0.54 (0.07 to 1.00)	0.0002	0.33 (-0.13 to 0.78)	0.04	0.21 (-0.25 to 0.67)	0.24
Teamwork	3.73 (0.11)	3.12 (0.14)	2.96 (0.14)	0.78 (0.45 to 1.11)	<.0001	0.62 (0.19 to 1.04)	<.0001	0.16 (-0.26 to 0.59)	0.35
Decision Making	3.59 (0.16)	3.13 (0.15)	2.72 (0.15)	0.87 (0.39 to 1.35)	<.0001	0.46 (-0.07 to 0.99)	0.0056	0.41 (-0.06 to 0.88)	0.02
Element									
Understanding of team members role	3.15 (0.14)	2.84 (0.16)	2.84 (0.16)	0.30 (-0.13 to 0.74)	0.10	0.31 (-0.20 to 0.82)	0.09	-0.00 (-0.49 to 0.48)	1.00
Attentiveness	3.53 (0.14)	3.31 (0.15)	3.15 (0.15)	0.38 (-0.05 to 0.82)	0.01	0.22 (-0.28 to 0.71)	0.24	0.17 (-0.30 to 0.63)	0.43
Professional modesty	3.60 (0.10)	3.14 (0.12)	3.16 (0.12)	0.44 (0.14 to 0.74)	0.0008	0.46 (0.12 to 0.81)	0.0004	-0.02 (-0.39 to 0.35)	0.98
Flexibility	3.48 (0.13)	3.01 (0.15)	2.80 (0.15)	0.68 (0.28 to 1.07)	<.0001	0.47 (-0.06 to 0.99)	0.0016	0.21 (-0.26 to 0.68)	0.26
Efficient use of team members	3.36 (0.13)	3.10 (0.16)	3.13 (0.16)	0.23 (-0.17 to 0.64)	0.24	0.26 (-0.29 to 0.81)	0.17	-0.03 (-0.51 to 0.46)	0.98
Uncertainty management	3.45 (0.19)	3.08 (0.18)	2.81 (0.18)	0.64 (0.06 to 1.21)	0.0001	0.37 (-0.15 to 0.89)	0.04	0.27 (-0.29 to 0.83)	0.19
Decision Analysis	3.37 (0.14)	2.94 (0.16)	2.69 (0.16)	0.68 (0.26 to 1.10)	<.0001	0.43 (-0.08 to 0.95)	0.009	0.25 (-0.24 to 0.74)	0.20
Leadership	3.34 (0.19)	3.03 (0.22)	2.70 (0.22)	0.64 (0.06 to 1.22)	0.0002	0.31 (-0.28 to 0.90)	0.10	0.33 (-0.34 to 1.00)	0.12
Prioritization	3.29 (0.20)	3.12 (0.19)	2.80 (0.19)	0.49 (-0.12 to 1.10)	0.0098	0.17 (-0.42 to 0.76)	0.13	0.32 (-0.24 to 0.88)	0.56

SD, standard deviation; Diff. Difference; CI, confidence interval

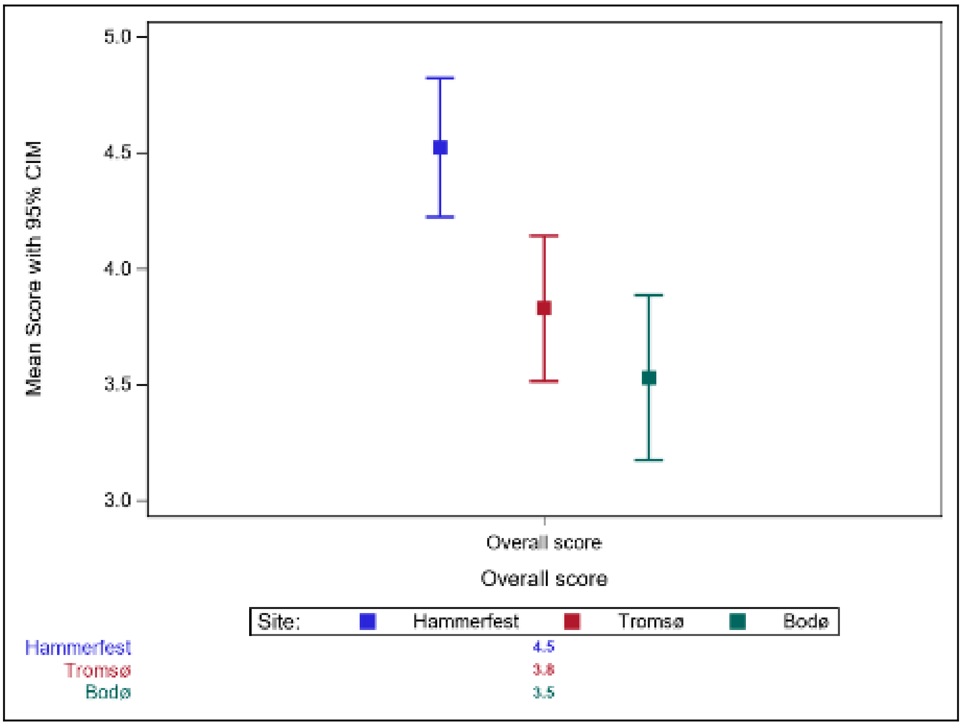


Figure 1 Overall score of NTS performance between the three training sites (CIM: Confidence interval of the mean)

NTS performance at the category level

The NTS performance levels of the students from Finnmark were significantly higher than those of the students from Bodø and Tromsø in all categories (Figure 2). We observed no significant difference between the NTS performance levels of the students from Bodø and Tromsø, except in terms of the category of “decision making”. The NTS performance levels of the students from Bodø were significantly lower than those of the students from Tromsø for this category. The categories of “Communication” and “Teamwork” exhibited the most significant differences between Finnmark and the other training sites. These results were also correlated with the overall score.

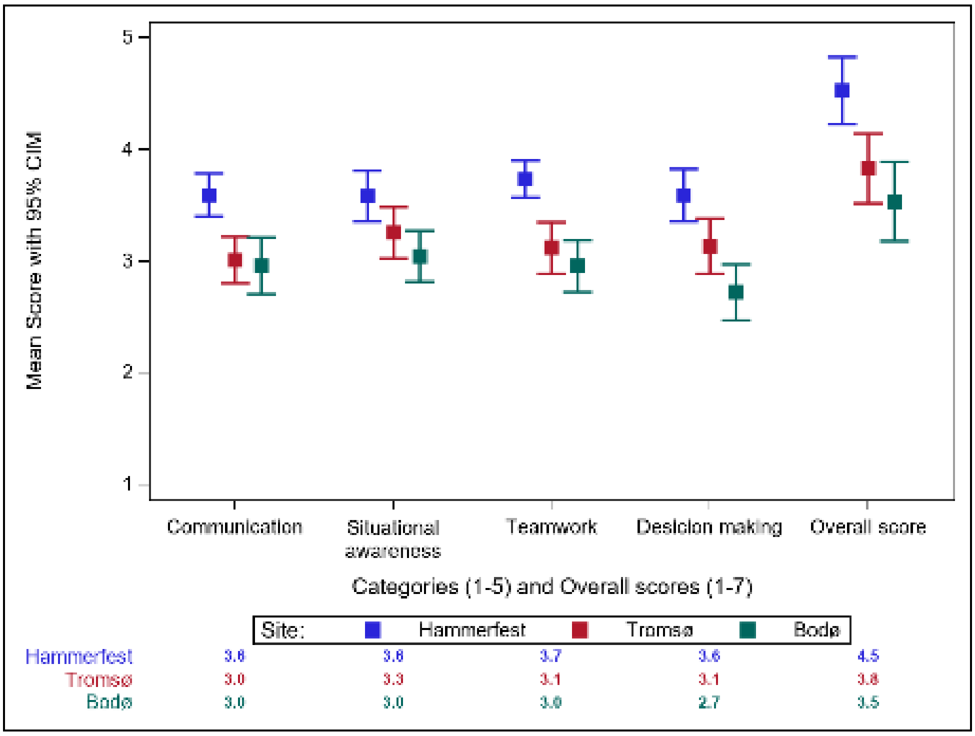


Figure 2 Category score of NTS performance between the three training sites (CIM: Confidence interval of the mean)

NTS performance when comparing elements

The NTS performance levels of the students from Finnmark were significantly higher than those of their peers in Bodø in all elements (Table 2), except the following elements: “Establish mutual understanding,” “Understanding of team members role” and “Efficient use of team members” (Figure 3). The element “Patient communication” exhibited the most significant difference. The scores for “Patient communication,” “Professional modesty,” “Flexibility,” “Uncertainty management,” and “Decision analysis” of the students from Finnmark were significantly higher than those of the students from Tromsø. We observed no significant difference between the students from Tromsø and Bodø in terms of these elements.

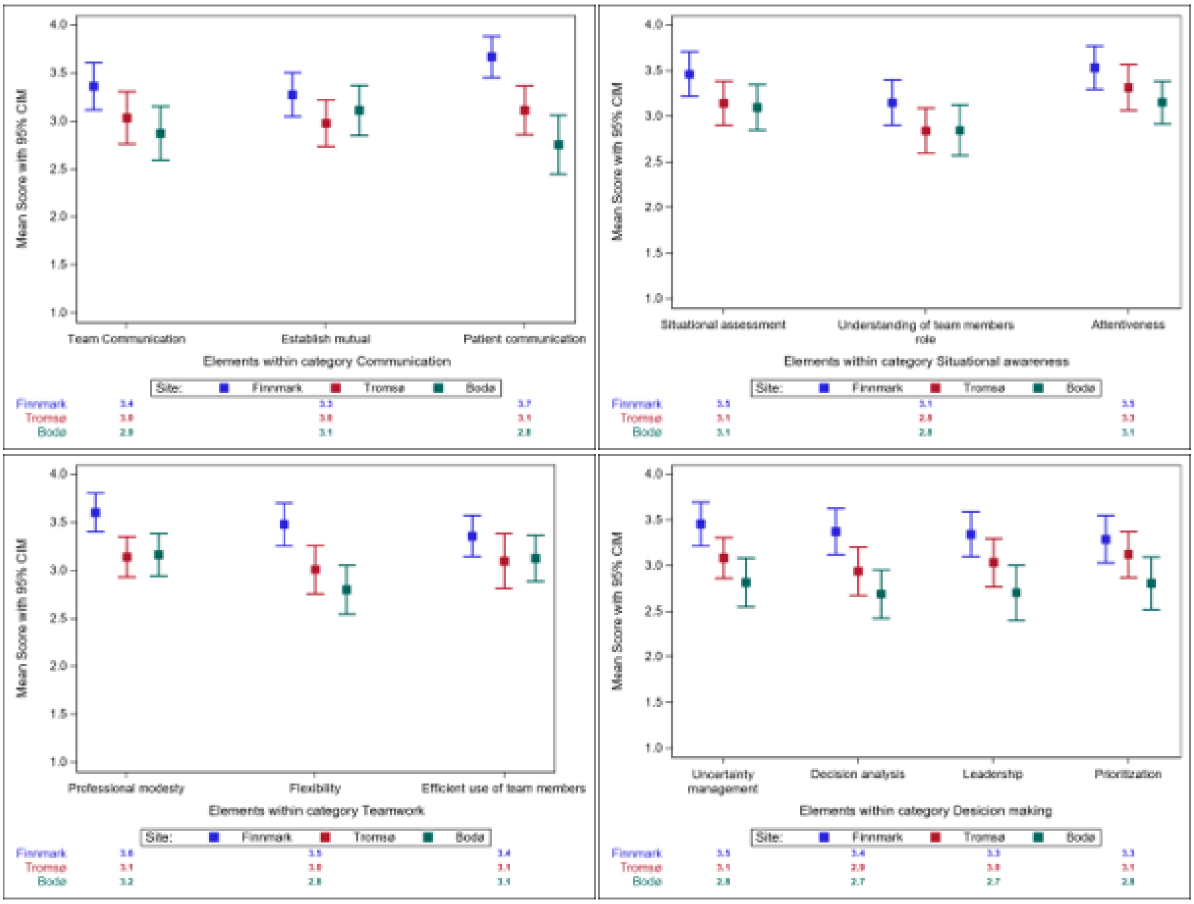


Figure 3 Element score of NTS performance between the three training sites (CIM: Confidence interval of the mean)

Discussion

The present study demonstrated that the overall and category-level NTS performance levels of the students from Finnmark were significantly higher than those of the students from Bodø and Tromsø. Moreover, we observed no significant difference between the NTS performance levels of the students from Tromsø and Bodø, except in the category "Decision making." The scores of the students from Bodø for this category were significantly lower than those of the students from Tromsø. The NTS performance level of the students from Finnmark was significantly higher for the elements "Patient communication," "Professional modesty," "Flexibility," "Uncertainty management," and "Decision analysis." This finding of consistent results, with no differences between the NTS performance of the students from Bodø and Tromsø, and significantly better NTS performance of the students from Finnmark supports the internal consistency of the tool. The present study is novel in that no previous study has evaluated the NTS performance of Norwegian medical students receiving decentralized medical education. Hence, the results of previous studies cannot be compared with our results.

UiT aimed to use multi-professional team training to enhance the NTS when developing the Finnmark model. The findings of the present study, a follow-up study conducted to evaluate its effect, indicate that the NTS performance levels of the students from Finnmark were superior. However, this finding can be attributed to several reasons.

First, selection bias may have influenced our results. All students in Finnmark and Bodø actively elected to pursue decentralized medical education during the fifth and sixth years of study. The students electing to go to rural areas may possess some distinct features prior to their choice that are associated with higher NTS levels. Nevertheless, the substantial difference observed indicates that is unlikely that the students possessed such levels of NTS performance when commencing the fifth year of study in Finnmark.

The student groups in Finnmark are smaller than those in Bodø and Tromsø. Small group teaching optimizes learning in healthcare.(25) The knowledge of the students increases when they can build their understanding with their peers.(25) Small group teaching also promotes team-building skills.(25) The student group in Finnmark is smaller; consequently, the lectures also contained small groups of students. Small-group teaching and small-group lectures are not equivalent.(25) However, the learning experience is correlated with the engagement of the students. Thus, small group lectures may facilitate higher learning outcomes than the bigger groups in Tromsø and Bodø. Furthermore, smaller groups may facilitate active participation, “face-to-face” contact between participants, and purposeful activities, which are the three key elements for small group teaching.(25)

More of the learning occurs in general practice in Finnmark, with general practitioners acting as teachers. The students received one-on-one half-day training in gynecology from an experienced general practitioner. The students also participated in general practice group workshops. Three students and one experienced general practitioner consulted with patients requiring a medical interpreter. The patient had regular appointments, and one of the students regularly consulted with a medical interpreter via telephone. The remaining students and general practitioners observed the consultation and participated in discussions. All students conducted one consultation by themselves. These new teaching models may have affected NTS positively as students are more engaged and active in the process and will receive feedback on different level that might help them with NTS. Those effects warrant further studies.

Another noteworthy difference is that the training site located in Finnmark is a small local hospital, whereas the training sites located in Tromsø and Bodø are a large University hospital and a large regional hospital, respectively. The local hospital in Finnmark comprised more generalists, whereas regional and university hospitals comprised more branch specialists. Consequently, more generalists trained the students in Finnmark, which may have affected the results. Students are expected to be skilled professional generalists by the time they graduate from medical school with the ability to become lifelong learners. The outcomes may be affected if education is particularly narrow or specialized. There is also a possibility that the teachers and the whole community in rural Finnmark places higher value on NTS. Further studies should aim to clarify these findings.

Limitations

The scenarios were standardized. However, they were not performed in an identical manner by the facilitators as they had different levels of training and different ways of performing their roles. Facilitators may have affected the ability of the students to perform at their highest level negatively and positively. Notably, several facilitators were involved at each institution to mitigate the influence of individual facilitators.

This study included eight students randomly selected from each training site. Ideally, all students should have been assessed; however, this was not possible owing to practical limitations. We obtained 64 scores for each element, category, and the global score of the NorMS-NTS at each training site as eight raters participated in this study. However, the generalizability of our findings remains unknown. Further studies must be conducted to validate these results and assess their applicability.

The NorMS-NTS is a novel assessment tool used to evaluate the NTS of Norwegian medical students. The process of collecting evidence for its validity is ongoing. Although not proven optimal for summative assessment, it is the only tool available to assess the NTS of Norwegian medical students. This may have affected the results. However, with one exception, we found no statistical difference between the NTS performance levels of the students from Bodø and Tromsø, which supports the reliability of the tool.

Conclusion

The NTS performance levels of the medical students in Finnmark were significantly higher than that of the students in Bodø and Tromsø. Further studies must explore the reasons for this discrepancy. However, our study demonstrated that rural decentralized medical education may yield better learning outcomes than standard education in large, centralized hospitals.

Acknowledgements

All authors contributed to the planning of this study. KP and TW recruited raters. KP did the videorecording of the students. HF and KP did the statistical analysis. HF, KP and TW did the interpretation of the work. KP drafted the manuscript. TW, PD and DM were major contributors to revising the manuscript. All authors contributed to the final version and gave approval for it to be published.

Competing interests

Dieckmann holds a professorship with the University of Stavanger that was established by an unconditional grant by the Laerdal foundation to the university and is today financed by the University itself. He is in the leadership of the EuSim group, providing faculty development courses in healthcare simulation.

The other authors declare that they have no competing interests.

Funding

This work was supported by the Northern Norway Regional Health Authority (Ref: HNF1368-17).

References

1. Aaraas IJ, Halvorsen PA, Aasland OG. Supply of doctors to a rural region: occupations of Tromsø medical graduates 1979-2012. Med Teach. 2015;37(12):1078-82.
2. Bertelsen T. Hvor kommer lægene fra og hvor blir de av. Tidsskr Norske Lægeforening. 1963;83:861-70.
3. Olsen IP, Lillebo B, Ofstad EH, Østlyngen A, Lappegård KT. Time to expand decentralisation of medical education in Norway. Tidsskr Norske Lægeforening. 2021;141(11).
4. Westlie Å L, Gaski M, Abelsen B, Grimstad H, Ofstad EH. Doctors educated in Bodø - who are they and where do they work? Tidsskr Nor Laegeforen. 2022;142(1).
5. Flin R OCP. Safety at the Sharp End: A Guide to Non-technical Skills: Ashgate; 2008.
6. Manser T. Teamwork and patient safety in dynamic domains of healthcare: a review of the literature. Acta anaesthesiologica Scandinavica. 2009;53(2):143-51.
7. Fletcher G, McGeorge P, Flin R, RJ G, Maran N. The role of non-technical skills in anaesthesia: a review of current literature. British journal of anaesthesia. 2002;88(3):418-29.
8. Fletcher G, Flin R, McGeorge P, et.al. Rating non-technical skills: developing a behavioural marker system for use in anaesthesia. Cognition, Technology & Work. 2004;6(3):165-71.
9. Prineas S, Mosier K, Mirko C, Guicciardi S. Non-technical skills in healthcare. In: Donaldson L, Ricciardi W, Sheridan S, Tartaglia R, editors. Textbook of Patient Safety and Clinical Risk Management. Cham: Springer International Publishing; 2021. p. 413-34.
10. Nestel D, Walker K, Simon R, Aggarwal R, Andreatta P. Nontechnical skills: An inaccurate and unhelpful descriptor? Simul Healthc. 2011;6(1):2-3.
11. Hagemann V, Herbstreit F, Kehren C, Chittamadathil J, Wolfertz S, Dirkmann D, et al. Does teaching non-technical skills to medical students improve those skills and simulated patient outcome? Int J Med Educ. 2017;8:101-13.
12. Neily J, Mills PD, Young-Xu Y, Carney BT, West P, Berger DH, et al. Association between implementation of a medical team training program and surgical mortality. Jama. 2010;304(15):1693-700.
13. Gordon M, Darbyshire D, Baker P. Non-technical skills training to enhance patient safety: a systematic review. Med Educ. 2012;46(11):1042-54.
14. Flin R, Martin L. Behavioral Markers for Crew Resource Management: A Review of Current Practice. The International Journal of Aviation Psychology. 2001;11(1):95-118.
15. Yule S, Flin R, Paterson-Brown S, Maran N, Rowley D. Development of a rating system for surgeons' non-technical skills. Med Educ. 2006;40(11):1098-104.
16. Prydz K, Dieckmann P, Musson D, Wisborg T. The development of a tool to assess medical students' non-technical skills – The Norwegian medical students' non-technical skills (NorMS-NTS). Med Teach. 2022:1-8.

17. Wang X, Kattan MW. Cohort Studies: Design, Analysis, and Reporting. Chest. 2020;158(1, Supplement):S72-S8.
18. Hospital SU. InterRegSim [Available from: https://www.helse-stavanger.no/interregsim.
19. Wisborg T, Brattebø G, Brattebø J, et.al. Training multiprofessional trauma teams in Norwegian hospitals using simple and low cost local simulations. Educ Health (Abingdon). 2006;19(1):85-95.
20. Kadam P, Bhalerao S. Sample size calculation. Int J Ayurveda Res. 2010;1(1):55-7.
21. Prydz K, Dieckmann P, Fagertun H, Musson D, Wisborg T. Collecting evidence of validity for an assessment tool for Norwegian medical students' non-technical skills (NorMS-NTS): usability and reliability when used by novice raters. BMC Med Educ. 2023;23(1):865.
22. Andrade C. Sample size and its importance in research. Indian J Psychol Med. 2020;42(1):102-3.
23. Montgomery DC. Design and analysis of experiments. 8th ed. Hoboken, N.J: Wiley; 2013.
24. Stoll A. Post hoc tests: Tukey honestly significant difference test. The SAGE Encyclopedia of Communication Research Methods 4: SAGE Publications, Inc; 2017.
25. Burgess A, van Diggele C, Roberts C, Mellis C. Facilitating small group learning in the health professions. BMC Med Educ. 2020;20(Suppl 2):457-.

Appendix

Appendix 1 – Approvals

Torben Wisborg
Hammerfest sykehus
9600 HAMMERFEST

Vår dato: 14.12.2017 | Vår ref: 57474 / 3 / HIT | Deres dato: | Deres ref:

Vurdering fra NSD Personvernombudet for forskning § 31

Personvernombudet for forskning viser til meldeskjema mottatt 28.11.2017 for prosjektet:

57474	*From student to professional – mastering the necessary non-technical skills? Assessment tools and effects of multi-professional simulation training*
Behandlingsansvarlig	*Finnmarkssykehuset, ved institusjonens øverste leder*
Daglig ansvarlig	*Torben Wisborg*

Vurdering

Etter gjennomgang av opplysningene i meldeskjemaet og øvrig dokumentasjon finner vi at prosjektet er meldepliktig og at personopplysningene som blir samlet inn i dette prosjektet er regulert av personopplysningsloven § 31. På den neste siden er vår vurdering av prosjektopplegget slik det er meldt til oss. Du kan nå gå i gang med å behandle personopplysninger.

Vilkår for vår anbefaling

Vår anbefaling forutsetter at du gjennomfører prosjektet i tråd med:

- opplysningene gitt i meldeskjemaet og øvrig dokumentasjon
- vår prosjektvurdering, se side 2
- eventuell korrespondanse med oss

Vi forutsetter at du ikke innhenter sensitive personopplysninger.

Meld fra hvis du gjør vesentlige endringer i prosjektet

Dersom prosjektet endrer seg, kan det være nødvendig å sende inn endringsmelding. På våre nettsider finner du svar på hvilke endringer du må melde, samt endringsskjema.

Opplysninger om prosjektet blir lagt ut på våre nettsider og i Meldingsarkivet

Vi har lagt ut opplysninger om prosjektet på nettsidene våre. Alle våre institusjoner har også tilgang til egne prosjekter i Meldingsarkivet.

Vi tar kontakt om status for behandling av personopplysninger ved prosjektslutt

Ved prosjektslutt 14.08.2023 vil vi ta kontakt for å avklare status for behandlingen av personopplysninger.

Se våre nettsider eller ta kontakt dersom du har spørsmål. Vi ønsker lykke til med prosjektet!

Dag Kiberg

Hildur Thorarensen

Kontaktperson: Hildur Thorarensen tlf: 55 58 26 54 / hildur.thorarensen@nsd.no

Vedlegg: Prosjektvurdering

Personvernombudet for forskning

Prosjektvurdering - Kommentar

Prosjektnr: 57474

Du har opplyst i meldeskjema at utvalget vil motta skriftlig informasjon om prosjektet, og samtykke skriftlig til å delta. Vår vurdering er at informasjonsskrivet til utvalget er mangelfullt utformet, og vi ber deg om å endre/tilføye følgende:
- det må fremgå at prosjektslutt er i 2023, og at data så vil bli lagret i inntil tre år etter dette for mulige oppfølgingsstudier, slik du har oppgitt i meldeskjema. Det bør også fremkomme om det kun er prosjektleder som vil bruke materialet på nytt, eller om det er snakk om bruk av data til andre forskningsprosjekter e.l.

Vi ber deg om å sende det reviderte informasjonsskrivet til personvernombudet@nsd.no Husk å oppgi prosjektnummer. Prosjektet kan deretter starte.

Du har opplyst i meldeskjema at TSD vil være databehandler i prosjektet etter avtale mellom Finnmarkssykehuset og TSD. Personvernombudet forutsetter at du behandler alle data i tråd med Finnmarkssykehuset sine retningslinjer for datahåndtering og informasjonssikkerhet.

Prosjektslutt er oppgitt til 14.08.2023. Det fremgår av meldeskjema/informasjonsskriv at du/dere vil lagre datamaterialet med personopplysninger frem til 01.08.2026 for oppfølgingsstudier/ny forskning og så anonymiseres. Vi gjør oppmerksom på at eventuell ny bruk av datamaterialet vil kunne kreve ny melding til ombudet.

Appendix 2 – NorMS-NTS tool

HELSE NORD | DAVVIDEARVVAŠVUOHTA VARRESVUOHTA NUORTTA HEALSOE NOERHTE

NorMS-NTS

Student nr: __

Kategori**	Kategori vurdering*	Elementer**	Element vurdering*	Tilbakemelding
Kommunikasjon		Kommunikasjon med team		
		Etablere felles forståelse		
		Kommunikasjon med pasient		
Situasjonsbevissthet		Sette seg inn i situasjonen		
		Forstå ulike roller i teamet		
		Oppmerksomhet		
Samarbeidsevne		Ydmykhet		
		Fleksibilitet		
		Bruke teamets ressurser		
Beslutningstaking		Gjøre gode valg		
		Håndtere usikkerhet		
		Lederskap		
		Prioritering		

Generelle kommentarer:

____

____

____

*N, Ikke observert. 1, Langt under gjennomsnittet; 2, under gjennomsnittet; 3, akseptabel; 4, over gjennomsnittet; 5, langt over gjennomsnittet..

** Innad i teamet om ikke annet er spesifisert.

Samlet vurdering (Sett ring rundt)**:**

Svært dårlig 1 – 2 – 3 – 4 – 5 – 6 – 7 Utmerket

Appendix 3 – NorMS-NTS manual

NorMS-NTS Håndbok

Norwegian **M**edical **S**tudents' **N**on-**T**echnical **S**kills

Et verktøy for tilbakemelding og vurdering av norske medisinstudenters ikke-tekniske ferdigheter

NorMS-NTS Versjon 1.0 Mai 2023

Kategori**	Kategori vurdering*	Elementer**	Element vurdering*	Tilbakemelding
Kommunikasjon		Kommunikasjon med team		
		Etablere felles forståelse		
		Kommunikasjon med pasient		
Situasjonsbevissthet		Sette seg inn i situasjonen		
		Forstå ulike roller i teamet		
		Oppmerksomhet		
Samarbeidsevne		Ydmykhet		
		Fleksibilitet		
		Bruke teamets ressurser		
Beslutningstaking		Gjøre gode valg		
		Håndtere usikkerhet		
		Lederskap		
		Prioritering		

For ytterligere informasjon eller spørsmål, ta gjerne kontakt:

Katrine Prydz, spesialist i allmennmedisin, fastlege. Katrine.prydz@gmail.com

Finnmarkssykehuset

Avdeling for fag, forskning og samhandling

Brenneriveien 19, 21.

9601 Hammerfest

https://finnmarkssykehuset.no/fag-og-forskning/lis

Forsidefoto: Finnmarksmodellen UiT

Hva er ikke-tekniske ferdigheter?

Ikke-tekniske ferdigheter er ferdigheter som kommunikasjon, beslutningstaking, lederskap, evne til å håndtere stress og usikkerhet. I 70% av uheldige hendelser på sykehus er ikke-tekniske ferdigheter vist å være en del av årsaken. Ved å øke medisinstudenters ikke-tekniske ferdigheter kan man øke nyutdannede legers ikke-tekniske ferdigheter, og dermed redusere risikoen for uheldige hendelser og øke pasientsikkerheten. Ikke-tekniske ferdigheter kan trenes opp. Det er derfor viktig at studentene får god opplæring i dette under studiet.

For å kunne vurdere studentenes ikke-tekniske ferdigheter er det nyttig å ha et verktøy. Det finnes verktøy for en del helsepersonellgrupper, men vi har ikke hatt verktøy for norske medisinstudenter. Studier har vist at det er nødvendig med egne verktøy for ulike land og ulike helsepersonellgrupper. På bakgrunn av det utviklet vi NorMS-NTS, som er et verktøy for å vurdere norske medisinstudenters ikke-tekniske ferdigheter.

Verktøyet skal kunne brukes for å gi tilbakemelding til studentene under studiet, slik at de kan jobbe videre med sine ferdigheter. Dette vil være det viktigste for å øke studentenes ikke-tekniske ferdigheter. Det er også ønskelig at det kan brukes til å evaluere undervisningen som gis, og at man kan sammenligne ferdigheter på gruppenivå. I tillegg vil bruk av verktøyet gi et økt fokus på viktigheten av ikke-tekniske ferdigheter.

Verktøy for ikke tekniske ferdigheter valideres for ulik bruk og ulik setting. Denne prosessen er i gang. Det er i første omgang validert for bruk av leger som har lite opplæring i bruk av verktøyet. Dette for at veiledere i praksis skal kunne bruke det som en vurdering for læring når de veileder studenter, altså formativ vurdering. Vi valgte denne valideringen først, for at verktøyet skulle være klart til bruk for alle som ønsker i en travel klinisk hverdag. Verktøyet kan per i dag ikke brukes for summativ vurdering, for eksempel til karaktersetting.

Verktøyet skal være i kontinuerlig utvikling og tilpasning. Den nyeste versjonen vil til enhver tid ligge på vår nettside: https://finnmarkssykehuset.no/fag-og-forskning/lis. Vi ønsker utstrakt bruk av verktøyet, slik at verktøyet kan valideres for ulike settinger. Ta gjerne kontakt dersom du ønsker å bruke verktøyet til forskning.

Kommunikasjon

Evne til å utveksle informasjon på en god måte.

Kommunikasjon med team

God atferd:

- Closed loop communication
- Bruker navn på teammedlemmer
- Snakker direkte til teammedlemmer
- Tydelige beskjeder og tilbakemeldinger
- Aktiv lytter

Dårlig atferd:

- Delegerer oppgaver uten å gi de til bestemte teammedlemmer, og uten å få bekreftelse.
- Uklare beskjeder
- Lytter ikke til teammedlemmer

Etablere felles forståelse

God atferd:

- Oppsummere situasjon og videre plan med hele teamet
- Forsikre seg om at alle forstår hvor akutt det er

Dårlig atferd:

- Ingen felles oppsummering
- Diskuterer litt med ulike kolleger

Kommunikasjon med pasient

God atferd:

- Presentere seg for pasienten
- Få pasienten til å fortelle hva som har skjedd
- Informere pasienten om vurdering og videre plan
- Tar hensyn til at pasienten er tilstede når det kommuniseres med team

Dårlig atferd:

- Presenterer seg ikke for pasienten
- Snakker ikke med pasienten
- Informerer ikke pasienten

SITUASJONSBEVISSTHET

Evne til å ha oversikt over situasjonen og følge utviklingen.

Sette seg inn i situasjonen

God atferd:

- Innhenter relevant informasjon fra teamet og pasienten for å skaffe seg oversikt over situasjonen
- Stiller spørsmål ved eventuelle uklarheter
- Observerer pasienten og teamets arbeid

Dårlig atferd:

- Ustrukturert innhenting av informasjon
- Etterspør ikke informasjon
- Går i gang med egne undersøkelser uten å vite hva som er gjort

Forstå ulike roller i teamet

God atferd:

- Introduserer seg med navn og yrke
- Avklarer teammedlemmers navn og yrke
- Er bevisst på teammedlemmers varierende kompetanse og erfaring

Dårlig atferd:

- Introduserer seg ikke
- Avklarer ikke teammedlemmers kompetanse og rolle

Oppmerksomhet

God atferd:

- Får med seg det som sies og gjøres i teamet
- Får med seg endringer hos pasienten

Dårlig atferd:

- Får ikke med seg det som sies og gjøres i teamet
- Får ikke med seg endringer hos pasienten

SAMARBEIDSEVNE

Evne til å bruke hele teamet på en god måte.

Ydmykhet

God atferd:

- Bevisst egne begrensninger
- Åpen for innspill fra team

Dårlig atferd:

- Overstyrer teammedlemmer
- Overser innspill fra team

Fleksibilitet

God atferd:

- Tilpasser seg raskt ved endringer hos pasienten eller i teamet
- Kommunikasjon og atferd tilpasses situasjonen

Dårlig atferd:

- Fortsetter som planlagt, tross endringer
- Tilpasser ikke kommunikasjonen til pasienten og situasjonen.

Bruke teamets ressurser

God atferd:

- Delegerer oppgaver til hele teamet
- Ber om innspill

Dårlig atferd:

- Gjør det meste selv
- Delegerer kun til en kollega

BESLUTNINGSTAKING

Evne til å fatte gode beslutninger til rett tid.

Gjøre gode valg

God atferd:

- Klargjør alternativer
- Vurderer og diskuterer problemstilling med team
- Ber om innspill fra kolleger
- Diskuterer tiltak
- Sjekker eller diskuterer retningslinjer

Dårlig atferd:

- Ingen tydelig vurdering av problemstilling
- Diskuterer ikke med team
- Avviser innspill
- Vurderer ikke retningslinjer
- Tar forhastede beslutninger
- Ser kun et alternativ

Håndtere usikkerhet

God atferd:

- Er åpen om egen usikkerhet
- Vurderer usikkerhet rundt prøvesvar og undersøkelser
- Er bevisst på at andre teammedlemmer kan ha mer erfaring
- Begrenser undersøkelser og tiltak til det nødvendige
- Tar beslutninger
- Konfererer ved behov
- Revurderer ved behov

Dårlig atferd:

- Forsøker å skjule usikkerhet
- Er skråsikker
- Stoler i hovedsak på egne vurderinger
- Gjør flere undersøkelser enn nødvendig
- Gjentar undersøkelser
- Overbehandler
- Unngår beslutninger
- Ingen revurdering

Lederskap

God atferd:

- Kommuniserer beslutninger tydelig til teamet
- Oppdaterer teamet på utvikling
- Gir tilbakemeldinger til teammedlemmer

Dårlig atferd:

- Kommuniserer ikke beslutninger
- Oppdaterer ikke teamet på utvikling

Prioritering

God atferd:

- Prioriterer oppgaver og tiltak etter behov og hastegrad
- Bruker ABCDE eller annen systematisk tilnærming

Dårlig atferd:

- Oppgaver og tiltak prioriteres ikke
- Bruker ikke systematisk tilnærming