Delft University of Technology

Document Version
Final published version

Licence
CC BY-NC-ND

Citation (APA)
Kes, L., Specht, M. M., Abbink, D., & Saunders-Smits, G. (2026). Critical Success Factors for Programmatic Assessment of Communication and Teamwork in Aircraft Maintenance Engineering Education. *SEFI Journal of Engineering Education Advancement*, *3*(2), 78-112. https://doi.org/10.62492/sefijeea.v3i2.67

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

SEFI Journal of Engineering Education Advancement
Vol. 3, No. 2, 2026, pp. 78-112, ISSN 3006-6301
DOI : 10.62492/sefijeea.v3i2.67

Critical Success Factors for Programmatic Assessment of Communication and Teamwork in Aircraft Maintenance Engineering Education

Lydia Kes[a,c,1], Marcus M. Specht [a, b], David A. Abbink[a] and Gillian N. Saunders-Smits[a]

[a] Delft University of Technology, Delft, The Netherlands, [b] Fern Universität Hagen, Hagen, Germany, [c] MBO College Airport, Hoofddorp, The Netherlands

ABSTRACT

The aviation sector is currently facing a growing shortage of aircraft maintenance engineers, further complicated by outdated training and assessment regulations focusing on theory and time requirements rather than competencies. In response, the industry is moving towards Competency-Based Training and Assessment, which introduces a significant shift by including transversal competencies that hold strong attitudinal elements requiring assessment through direct observation of performance. This study seeks to identify the critical success factors for training and assessing these transversal competencies within aircraft maintenance vocational education and training, leveraging principles of programmatic assessment.

A five-month action research study was conducted involving 21 students and three instructors, utilizing a previously developed training and assessment programme validated by industry stakeholders. The findings demonstrate that a programmatic assessment approach, integrating peer-, self-, and instructor- assessments within individual assessment portfolios, is perceived as effective to develop and evaluate competence. Students report enhanced reflection, self-awareness, and competency performance. Key factors for success included the adoption of a coaching approach, the use of simulation and role-play, and the implementation of iterative assessments with formative feedback in supportive environments. Additionally, instructors benefited from applying the ORCA (Observe, Record, Classify, Assess) method and undertaking further training in holistic group assessment.

KEYWORDS

Competency-Based Training and Assessment, Transversal Competencies, Aircraft Maintenance Education, Programmatic Assessment.

PUBLICATION

Submitted:
10th June 2026

Accepted after revision:
9th August 2026

Introduction

The aviation industry is becoming increasingly complex, posing more and more challenging engineering work around technological advancements that enhance safety and operations (Amin et al., 2022; Ichou & Veress, 2023). Yet the shortage of engineers has been mounting alarmingly in the post-

[1] Corresponding Author : Lydia Kes – L.Kes@tudelft.nl

This work is licensed under Attribution-NonCommercial-NoDerivatives 4.0 International. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/

COVID-19 period (Boeing, 2022; Costanza et al., 2022). Consequently, the aviation industry urgently needs to rethink how to increase labour productivity on the one hand, and train more, and more competent, technical personnel.

The focus here is placed on increasing the efficiency of training and assessment programmes towards this societal challenge. Existing European regulations for aircraft maintenance training have emphasised theoretical instruction and time-based assessments for decades (European Union Aviation Safety Agency (EASA), 2022). However, this approach is no longer suitable to prepare future engineers for the complex and fast developing industry (Kearns et al., 2016; Zylawski & Ma, 2023). Therefore, around the world, aircraft maintenance education and training organisations are exploring a shift towards Competency-Based Training and Assessment (CBTA) considering training quality, efficiency and assessment of complex task performance. For instance, engineers are required to diagnose and rectify an intermittent system fault under time pressure, integrating technical troubleshooting, documentation, and regulatory knowledge while coordinating with multiple stakeholders. It involves continuous communication with team members, maintenance control, and flight crew to align diagnostic findings and operational constraints. Final airworthiness certification depends on both accurate technical judgement and effective communication to ensure safe release to service.

To facilitate the transition to CBTA, the International Civil Aviation Organisation (ICAO) has introduced a competency framework for aircraft maintenance, encompassing both technical and non-technical, transversal competencies (TVCs) (ICAO, 2020). TVCs are related to responsibility and initiative, teamwork, communication, coordination, problem solving and critical thinking (Villardón-Gallego et al., 2020; Winch & Addis, 2022). Examples of TVCs from ICAO's competency framework are work management, teamwork and communication (Kes, Saunders-Smits, et al., 2026). While CBTA has not yet been established as a new global standard, its anticipated adoption represents a significant paradigm shift in aircraft maintenance education and training, most notably the integration of TVCs, and the assessment of behaviour.

ICAO defines competency as:

> A dimension of human performance that is used to reliably predict successful performance on the job. A competency is manifested and observed through behaviours that mobilize the relevant knowledge, skills and attitudes to carry out activities or tasks under specified conditions (ICAO, 2020).

These two sentences represent a major shift in assessment within aircraft maintenance: from standardised testing to reliably predicting performance by observing behaviour, and considering knowledge, skills, and attitudes holistically. Currently, the licensing process for aircraft maintenance technicians in Europe primarily relies on multiple-choice knowledge exams and time-based experience assessments. While competencies and attitudes are acknowledged, they are not systematically assessed (EASA, 2022).

Education research shows that competencies can only be assessed through the observation of performance and behaviour (Gulikers et al., 2018; McClelland, 1973; Spencer & Spencer, 1993; Van der Vleuten, 1996). In other words, achieving high scores on multiple-choice examinations and fulfilling the required experience hours does not necessarily indicate true competency. To accurately determine a student's competence, it is essential to evaluate their performance and behaviour in a context that is relevant to the field (Gulikers et al., 2007; Norcini et al., 2011), encompassing not only knowledge and skills but also performance and attitudes (Schuwirth & Van der Vleuten, 2011). While knowledge and skills are relatively straightforward to observe, knowledge representing what the student knows and understands as an engineer, and skills reflecting what the student can do, the

assessment of attitudes is more complex. Attitudes pertain to the type of engineer the student becomes, encompassing aspects such as motivation, personal traits, and values. These attitudinal components are inherently more challenging to develop and assess (Al Mamun, 2012; Spencer & Spencer, 1993). The attitudinal aspect of competency is most evident in the TVCs.

Subjects such as communication, teamwork, stress management, assertiveness and awareness are far from new in aircraft maintenance training, most notably through the mandatory 'Human Factors' module that highlights their critical role in aviation safety (EASA, 2022). However, the prevailing challenge remains the reliance on traditional, knowledge-focused assessment methods. For example, to complete the module 'Human Factors', European students are expected to answer multiple-choice questions on topics like the anatomy of the inner ear, or the negative effects of drugs, rather than being assessed on their capacity to demonstrate practical and essential behaviours in the aviation context, such as effectively conducting a shift handover or reducing one's stress levels appropriately. This highlights the key challenge of this study: assessing TVCs by observing behaviour to reliably predict future performance.

Given the complex interplay between knowledge, skills, and attitudes, comprehensive evaluation of competencies requires programmatic assessment using various methods over a longer period of time, with multiple assessors, in real-world settings that closely reflect the actual professional environment (Baartman et al., 2006; Schuwirth & Van der Vleuten, 2011). Portfolios can be a useful tool to collect the assessment data (Krause et al., 2015; Lockyer et al., 2017) and the evidence collected can be used both formatively and summative (Prescott et al., 2002; Van Der Vleuten et al., 2018). Since assessments determine the learning process (Gerritsen-van Leeuwenkamp et al., 2019; Sluijsmans et al., 2008) the teaching and learning activities should align with the assessments (Biggs, 1996).

Aim of this Study

This study seeks to evaluate the perception and effectivity of a CBTA programme which integrates TVCs within the technical context of aircraft maintenance using a programmatic assessment approach by observing students and instructors during the course, analysing students' portfolios, and conducting focus group interviews with students and staff.

Before presenting the research questions and methodology, the next sections outline the study background, its place within a broader study on aircraft maintenance education and training, and the literature on assessing TVCs through assessments.

Background to this study

This study is part of a larger study and builds on previous research that identified the most crucial yet challenging TVCs and observable behaviours from ICAO's aircraft maintenance competency framework: Communication and Teamwork (Kes, Saunders-Smits, et al., 2026). A supplemental Delphi study with international experts defined performance levels, assessment context, knowledge, skills, attitudes, and training levels for these competencies and their related behaviours (Kes, Wesel, et al., 2025). Using these results, a training and assessment programme was developed and validated with stakeholder focus groups (Kes, Pietersma, et al., 2025, 2026). The programme targets the following competencies and observable behaviours (OBs) from the ICAO Framework:

- Competency 11 - Communication: *'Communicate effectively in all situations and ensure clear and common understanding'* with observable behaviour *11.4 'Maintains situational awareness when selecting method of communication, speaks clearly, accurately and concisely'* and

- Competency 8 – Teamwork: *'Operate safely and efficiently as a team member'* with observable behaviours 8.1 *'Fosters an atmosphere of open communication'* and 8.12 *'Anticipates and responds appropriately to the needs of others'* (ICAO, 2020).

Using Peer-Assessment to Assess TVCs

Van den Eertwegh et al. (2014) found that during communication assessments, medical students felt the process was unnatural when this competency was broken down into behavioural checkboxes, rather than recognizing the holistic and context-specific nature of communication. Because of its complex behavioural nature, communication should be observed directly in an authentic context over a longer period. Feedback should consist of narrative aggregated assessor data and align with the student's personal learning objectives when discussed during reflective dialogues (Van Den Eertwegh et al., 2014; Van der Vleuten et al., 2019).

Using anonymized peer assessments (PAs) emerge as useful in the training and assessment of both communication and teamwork. Using anonymized PAs potentially reveals elements of team dynamics that may not be observable from outside (Jones & Abdallah, 2013) and are considered more reliable since teachers simply cannot oversee all team dynamics and its corresponding interactions (Farland & Beck, 2019; Strom & Strom, 2011). Moreover, Strom & Strom (2011) indicate that using anonymized PAs for evaluating teamwork with students resembles how these competencies are often evaluated in the workplace using the 360° feedback approach rather than hierarchical management evaluations. When these peer-observations are compared with self-evaluations, students can learn to become self-critical, reflective, introspective and modify their behaviour based on peer-input. To obtain reliable peer evaluations, substantial interaction is recommended, where teams work together for several months and report based on consistently demonstrated behaviours. The reliability of observations increase with the number of observers (Strom & Strom, 2011).

Programmatic assessment, including instructor-, peer-, and self-assessments (SAs) of communication and teamwork during real technical maintenance tasks is considered valid and feasible by both the aircraft maintenance industry and students. This validity is contingent on the use of standardized rubrics and instructors who are experienced or trained in behavioural assessment (Kes, Pietersma, et al., 2025). In the next sections, literature on PA and SA in education will be further discussed.

Existing literature predominantly examines PAs and SAs within higher education settings, focusing on written tasks or product design. There is limited focus on vocational education. In particular, research addressing performance and behaviour assessments (beyond presentations) in vocational contexts remains scarce. However, PAs can be used to evaluate behaviours, like how students learn or interact socially (Topping, 2017).

Van Helden et al. (2023) distinguish several types of PAs; Type 1, peer review, involves students reviewing and providing feedback on each other's work, such as essays, reports, prototypes, or code. The recipient may or may not incorporate this feedback in later versions. Type 2, peer grading, has students assign grades to each other's work, like homework or essays, based on set criteria, typically without detailed feedback. Type 3, peer evaluation, focuses on students assessing group members' transversal competencies and contributions during collaborative assignments, including teamwork and effort.

Most studies on PA highlight students' preference for strict anonymity to reduce social pressure and fear of damaging relationships, leaving more room for honesty (Kearney, 2013; Resta & Lee, 2010; Seifert & Feliks, 2019). However, Tornwall (2018) discovered that anonymity is less appropriate in nursing education because it presents logistical challenges in clinical environments and reduces

chances for meaningful dialogue, accountability, and the development of supportive relationships. More studies underscore the positive effects of PAs on responsibility, communication, teamwork, metacognition, reflection and awareness (Lerchenfeldt et al., 2019; To & Panadero, 2019). Furthermore, it is suggested that students who engage in both PA and SA demonstrate a deeper understanding of standards and assessment criteria, and report higher levels of engagement with the learning process (Kearney, 2013). Additionally, PA and SA foster a more open, supportive, and energetic class climate, facilitating mutual help, practice opportunities, and clearer understanding of tasks (Ndoye, 2017). Altogether, it can be concluded that self- and peer assessments not only improve the competencies that require development but also, as a secondary effect, stimulate metacognitive attitudes such as awareness and reflection which enhances the learning process (Schön, 2013; Zimmerman, 2002), and that communication and teamwork in particular are fostered not only through the assessment content focusing on these abilities, but also through the assessment methods themselves.

Furthermore, however, several studies indicate that training is needed to optimize the positive effects of PAs and SAs. Students need to become familiar with assessment language, how to provide feedback, staying objective and interpret the assessment criteria correctly and consistently (Kearney, 2013; Lerchenfeldt et al., 2019; Resta & Lee, 2010). For that reason, it is important to start early in the programme with practicing these assessments (Tornwall, 2018). To practice PAs, Resta & Lee (2010) advise using a hypothetical case of a fictional team-member where students then evaluate the performance using a PA instrument.

Research Questions of this Study

As outlined in the preceding sections, the optimal approach to evaluating TVCs is through programmatic assessment conducted within authentic, profession-specific contexts. Given that communication and teamwork are recognized as paramount TVCs within the ICAO competency framework, it is essential to establish robust and practicable methods for their training and assessment. The literature suggests that the application of PA and SA, in combination with instructor assessment, constitutes an effective strategy for evaluating these competencies. Accordingly, the present study aims to investigate the implementation of a programmatic assessment model that incorporates peer-, self-, and instructor- assessments within realistic scenarios, with the objective of supporting the training and assessment of communication and teamwork in aircraft maintenance education. To explore whether this objective is realised in practice, the following research questions are articulated:

The main research question is: ***What are critical success factors for training and assessing TVCs in aircraft maintenance vocational education and training, using programmatic assessment principles?***

To address this overarching question, the following sub-questions are formulated:

1. *What is the perception of self-assessments, peer-assessments, and instructor- assessments within the programmatic assessment approach?*
2. *To what extent does the training and assessment programme facilitate the development of transversal competencies?*
3. *What are the key prerequisites for effectively implementing programmatic assessment of transversal competencies in competency-based aircraft maintenance training?*

Methods

To understand the critical success factors for training and assessing TVCs in aircraft maintenance vocational education and training (VET), Action Research (AR) was conducted in which a newly designed training and assessment programme was implemented.

AR can be defined as a small-scale intervention in real-world practice combined with the systematic examination, monitoring, and evaluation of its effects (Cohen et al., 2018). It integrates action and reflection in a structured process aimed at both understanding and improving practice. As a collaborative and transformative approach, AR emphasizes rigorous data collection, knowledge generation, and reflective inquiry, alongside deliberate actions that seek practical and context-specific solutions (Piggot-Irvine et al., 2015).

AR is particularly suited to educational contexts in which existing practices are reconsidered or replaced, for example when transitioning from traditional teaching approaches to new methods (Kemmis, 2009). It supports the evaluation of procedures, as well as the exploration of attitudes and values (Cohen et al., 2018), while simultaneously aiming to change practices, deepen understanding of those practices, and examine the conditions that shape them. In this sense, AR can be understood as a "practice-changing practice," especially relevant in situations involving paradigm shifts in education (Kemmis, 2009), such as moving towards programmatic assessment of competencies.

A key strength of AR lies in its capacity to bridge the gap between research and practice by integrating inquiry with action, rather than treating them as separate activities (Somekh, 1995). Researchers are not detached observers but active participants in the intervention, working collaboratively with practitioners and participants rather than conducting research *on* them (Merriam & Tisdell, 2016). Throughout the process, qualitative data are systematically collected as the intervention unfolds, allowing for iterative refinement and deeper understanding (Merriam & Tisdell, 2016; Piggot-Irvine et al., 2015; Somekh, 1995).

In educational settings, AR is inherently collaborative, often involving close cooperation between teachers and researchers (Cain, 2011; Cohen et al., 2018). Due to the complexity of human interactions and contextual influences, controlled experimental designs are generally not feasible. Instead, AR acknowledges the dynamic and relational nature of educational practice. Importantly, claims of improvement resulting from an intervention must be substantiated through consultation with participants, including students, who should be informed about and involved in the intended changes. In the current study, this requirement for participant substantiation was addressed through a confirmatory focus group session, which will be explained further in the section on data collection. While findings from AR are context-specific, they can inform and inspire similar practices in comparable settings (Cain, 2011).

AR is well suited to the present study as it examines the effects of a new educational approach in an authentic setting. Moreover, the research questions address stakeholders' perceptions, which aligns with AR's participatory nature, where both students and instructors contribute to data generation and reflection.

Research Setup

During this study, the designed and validated training programme was implemented in a training hangar with aircraft maintenance students. The research took place at MBO College Airport in Hoofddorp, The Netherlands, a VET institution near Schiphol Airport dedicated to aviation training. The Aircraft Maintenance Basic Training department serves approximately 500 students at European Qualification Framework (EQF) levels three and four, preparing them for careers in aircraft

maintenance. The department is led by two training managers: one oversees business operations, and the other is responsible for educational and curriculum-related matters.

A class of twenty-one fourth (final) -year students at EQF level four was selected. As they had already completed a 20-week internship in a real maintenance environment, this advanced stage of their studies was considered appropriate for integrating TVCs into the continued development of their aircraft inspection and troubleshooting skills. Not only are those key abilities essential for their future careers, but also during the validation of the programme, these activities were deemed effective by all stakeholders in fostering communication and teamwork in a meaningful context (Kes, Pietersma, et al., 2025, 2026). Since only a very small percentage of students enrolled in aircraft maintenance training is female, all participating students were male, with an average age of 20 years old. Both programme implementation and data collection took place in two stages, as outlined in figure 1.

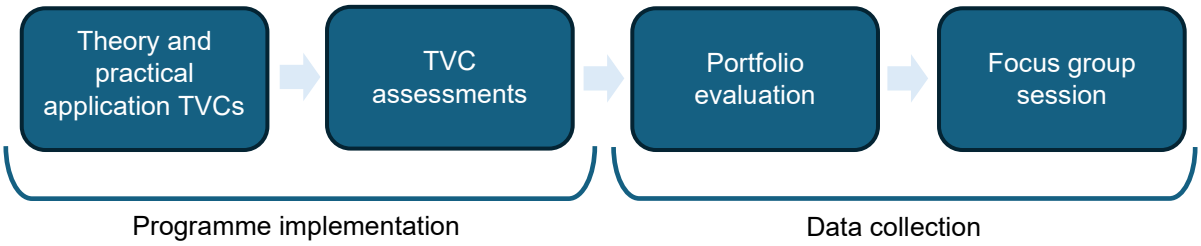


Figure 1: Schematic structure of research setup

The study involved a TVC training and assessment programme taught from September 2025 to February 2026 by two experienced aircraft maintenance instructors and assessors from the college, and one external instructor and assessor with over 40 years of experience in MRO training at a large Dutch airline. Students dedicated one full day a week to participate in this new TVC training and assessment programme. On the remaining days, they continued their regular curriculum, which included theoretical courses on aircraft structures, maintenance procedures, English, Mathematics, Physics, and an elective subject aligned with their interests. Details about the structure and content of the TVC training and assessment programme will be provided in the following sections.

Structure, Content and Timeline of the Programme

During a period of 15 weeks, students were trained on a weekly basis, both theoretically and practically, about communication and teamwork in aircraft maintenance. The total study can be divided into two stages. The first five weeks, stage 1, were solely focused on understanding the TVCs and how they are related to aviation safety and practicing them in an aviation context through a newly designed module. This module was designed by the first author in consultation with the two instructors from the college using aviation, teamwork, and communication literature, combined with social and communication skills training materials for VET students. The content was specifically adapted for aircraft maintenance. Stage 1 was followed by Stage 2, a ten-week period focused on assessment, during which peer-, self-, and instructor-assessments were collected in each student's portfolio. Both stages are described in more detail in the sections below.

ISSN 3006-6301, Vol. 3 (No. 2) 2026

Stage 1: Introduction to TVCs in Aircraft Maintenance Training

Each lesson day started with theoretical presentations, explanations, discussions and exercises about the TVCs. Students were explained how these competencies are recognized and they were introduced to the rubrics designed for evaluation. Also, videos were made for each OB so students could see what (in)competent behaviour looks like in an aircraft maintenance context. As suggested by Resta & Lee (2010), there were exercises included where students had to practice evaluating behaviour while watching videos or live role-play by the instructors. The TVC lessons were presented by the first author, and the three instructors provided supplemental practical examples from the aviation context to enhance understanding. Figure 2 provides a schematic display of how the module was built, illustrating progression from individual self-awareness (center), through interpersonal awareness and team interaction, to situational communication practices and, ultimately, the development of a shared team and safety culture (outer layer).

In addition, the technical tasks and activities the students were working on were discussed and evaluated, and, if needed, the instructors provided supplemental instruction about aircraft systems. After each theory lesson, all students went to the hangar to start the practical training, where all instructors adopted a coaching approach. All tasks and activities took place on a Boeing 737-200 in the training hangar. At the beginning of the course, students were randomly assigned to groups of three with whom they collaborated throughout the entire term. In total, seven groups of students were formed. This approach was implemented to ensure substantive interaction within teams over several months, thereby facilitating reliable peer evaluations of teamwork and communication (Strom & Strom, 2011). While working on their technical tasks, students practiced with the communication and teamwork OBs and received direct formative feedback on both their technical and transversal task performance. Figure 3 shows the training hangar where the study took place. After the practical training in the hangar, the students returned to the classroom to discuss the status of their maintenance tasks, reflect on their performance and to evaluate the day.



Figure 2: Model of training module

Note: Conceptual model developed by the authors



Figure 3: Training hangar with Boeing 737-200 used for this study

© Lydia Kes, 2026. All rights reserved. No reproduction or distribution without permission.

Stage 2: Programmatic Assessments

After five weeks, the students were sufficiently used to the setting, their new instructors and the focus on communication and teamwork. This is where the assessment phase was added. Students were extensively introduced to the different assessment types, and the rubrics were printed and placed in the hangar to remind students of the performance they were expected to focus on. Additionally, simulation and role-play were introduced. The simulation was situated around an A-check where critical maintenance tasks were conducted. Critical maintenance tasks are tasks directly affecting flight safety. They include work on the engines, landing gear and fuel system. Each group of students was assigned a different technical inspection task. A role-play component was incorporated to encourage targeted behaviours, with scenarios rotating each week. The lead instructor assumed multiple roles, such as a lazy supervisor, a foreign client with a specific cultural background, a national aviation authority inspector, or an inexperienced intern, among others. In this environment, where the authentic context was simulated as much as possible, each observable behaviour was assessed three times using three different assessment methods: SA, PA, and instructor-assessment (IA). The three instructors used the ORCA method: Observe, Record, Classify and Assess/Evaluate (International Air Transport Association, 2023) when the students performed their tasks. The assessments were based on the rubrics, where students had to evaluate for that day whether the performance for each element was rated as ‘Beginner’, ‘In Development’ or ‘Competent’ and they had to provide examples to support their rating. Tables 1, 2 and 3 show the SAs the students completed for the different OBs. The PAs and IAs were based on the same performance levels but written from the perspective of the evaluator. After each second round, formative feedback was given per group of students in the presence of, and guided by the instructors, as aircraft maintenance students consider this a socially safe setting to discuss PAs and to provide general feedback for improvement to each student (Kes, Pietersma, et al., 2025). During these sessions, students received feedback from both the instructors and their peers.

ISSN 3006-6301, Vol. 3 (No. 2) 2026

Table 1: Self-Assessment used for evaluating OB 11.4

Beginner	In Development	Competent
When communicating, I did not distinguish between information that was or was not safety critical. As a result, I communicated a lot of unnecessary information and/or omitted essential safety-related information.	When communicating, I somewhat distinguished between information that was or was not safety critical. I still communicated some unnecessary information and/or omitted some essential safety-related information.	I communicated only essential information that was important for safety. I did not communicate unnecessary information.
I usually or always assumed that the information I communicated was clear to the recipients without checking.	I sometimes assumed that the information I communicated was clear to the recipients without checking, or I checked only with a few recipients.	I never assumed that the information I communicated was clear to the recipients; I always checked with all recipients.
When I received verbal information, I did not check whether I understood it correctly.	When I received verbal information, I sometimes checked whether I understood it correctly, for example when I was unsure what the other person meant.	When I received verbal information, I always checked whether I understood it correctly by repeating it, possibly in different words.
I made many assumptions during communication without checking whether they were correct.	I made some assumptions during communication that should have been checked.	I did not make automatic assumptions; I always checked first whether they were correct.

Table 2: Self-Assessment used for evaluating OB 8.1

Beginner	In Development	Competent
I showed little to no interest in my fellow students or colleagues.	I showed some interest in my fellow students or colleagues	I showed full interest in my fellow students or colleagues.
I did not ask any questions, especially when I felt it was something I should know.	I asked a few questions, sometimes even when I felt it was something I should know.	I asked all the questions I wanted, including when I felt it was something I should know.
I did not discuss mistakes or errors with my fellow students/colleagues or instructor/supervisor.	I sometimes discussed mistakes or errors with my fellow students/colleagues or instructor/supervisor or limited this to one or a few team members.	I discussed all mistakes or errors I made and involved all team members, including the instructor/supervisor.
I involved few or no fellow students/colleagues in my communication, even when it was relevant to other team members.	I involved some team members in my communication, but not all those for whom it was relevant.	I involved all team members in my communication for whom it was relevant

Table 3: Self-Assessment used for evaluating OB 8.12

Beginner	In Development	Competent
I did not respond to non-verbal signals, such as body language from fellow students or colleagues, indicating they needed help.	I sometimes responded to non-verbal signals, such as body language from fellow students or colleagues, indicating they needed help.	I usually or always responded to non-verbal signals, such as body language from fellow students or colleagues, indicating they needed help.
I did not respond when my fellow student/colleague needed help while spending a very long time on a task.	I sometimes responded when my fellow student/colleague needed help while spending a very long time on a task.	I usually or always responded when my fellow student/colleague needed help while spending a very long time on a task.
I never offered my help proactively; it had to be requested.	I sometimes offered my help proactively, without being asked first.	I usually or always offered my help proactively, without being asked first.
When I offered help, I completely took over the task without explaining what to do, or I did so in a rushed/stressed manner.	When I offered help, I partly took over but still involved my fellow student/colleague somewhat in completing the task. I was still somewhat rushed/stressed.	When I offered help, I calmly guided my fellow student/colleague through the process, and we completed the task together.
I did not handle observed information about my fellow student/colleague discreetly and confidentially before, during, or after helping them.	I handled observed information about my fellow student/colleague partly discreetly and confidentially before, during, or after helping them.	I handled observed information about my fellow student/colleague fully discreetly and confidentially before, during, and after helping them.

The structure of the lesson days in stage 2 remained the same as in stage 1. During the theory in the morning the previous assessments were evaluated, and anonymized examples from the assessments were addressed plenary in the classroom to discuss which evaluations are well-substantiated and which are not. If necessary, certain rubric descriptions were further explained to ensure that they were interpreted correctly by the students. After the practical lesson in the hangar, the students returned to the classroom where they completed both the SAs and PAs. Most students used electronic assessment forms and uploaded them in their learning management system. Some students preferred using paper versions of the assessment forms. Figure 4 schematically shows the structure of the complete training programme.

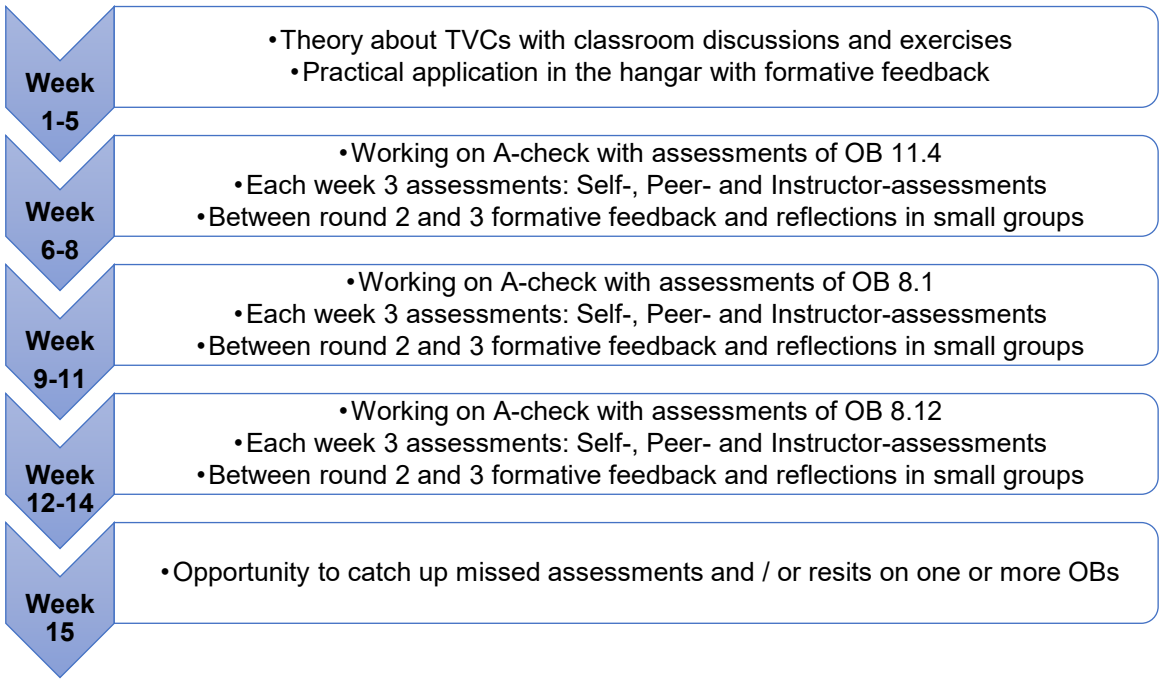


Figure 4: Schematic structure of TVC training and assessment programme (stages 1 and 2)

Data Collection

As outlined in the previous section, peer, self-, and instructor assessment data were collected in students' portfolios throughout the TVC training and assessment programme. These portfolios constituted one component of the data collection. Following week 15, the programme was evaluated through a portfolio assessment and a confirmatory focus group session. Each data source is discussed separately below.

Assessment Portfolios

The assessment portfolios were evaluated by the three instructors and the educational training manager of the Aircraft Maintenance Basic Training department of the college, in the presence of the first author. The educational training manager holds a master's degree in educational sciences and is highly familiar with the aircraft maintenance training curriculum. Her involvement was important to minimise potential bias during the discussion of the students' portfolios. During the portfolio evaluation, the current competency status of each student was examined. In this process, particular attention was paid to the justifications accompanying peer- and self-assessment ratings, as these provided insight into whether students were able to recognise, interpret, and evaluate observed behaviours in relation to the rubric criteria and associated performance levels. Consequently, the validity of students' justifications formed part of the overall judgement of competency status. Since peer- and self-assessment data may contribute to future competency judgements within CBTA, understanding the extent to which students can provide evidence-based and criterion-referenced ratings is important for evaluating the validity of these assessment processes.

The first author took notes on both the discussion process and the final evaluations of the students' portfolios.

Confirmatory Focus Group

To explore students' and instructors' perception, a confirmatory focus group session was held. Confirmatory focus groups are particularly useful to evaluate the efficacy, quality, feasibility and effectiveness of a new design after implementation (Gibson & Arnott, 2007; Tremblay et al., 2010). The students were selected to ensure a balanced representation of the class. Several students exhibited strong technical competencies and / or high performance on the TVCs, whereas others were continuing to develop their competencies. In total, nine students were invited for the session, of which seven students were able to participate, supplemented by two participating instructors. During the focus group session, students were asked to reflect on several aspects related to the training and assessment of the different competencies and OBs. Participants reflected on the impact of the training programme on their development and their experiences with assessing technical and transversal competencies during an A-check. Focus was placed on self- and peer-assessment, including confidence in judgement, bias, and critical reflection on communication and teamwork. They also evaluated key instructional elements and provided feedback on their motivation and overall learning experience.

The instructors were asked to share their experiences with instructor-led assessments, the role-play, and the integration of technical and behavioural instruction and feedback. To ensure that students felt comfortable expressing their views openly, it was explained at the outset that no hierarchy existed among focus group participants. It was further emphasised that neither the assessment outcomes nor their opinions about the training programme would have any negative impact on their regular scholar progression. The focus group session was led by the first author and recorded and transcribed using MS Teams. The transcripts were checked for accuracy and readability against the original recordings and then anonymized. They were subsequently coded using Atlas.ti, followed by further analysis and categorization to extract the main themes. The results will be presented in the next section.

Ethical permission was given by the Delft University of Technology's Human Research Ethical Committee for conducting the training and assessment programme for portfolio evaluation, and the post-programme focus group session (approval reference numbers 5843 and 6307 respectively).

Results

As described in the Methods section, data were collected across multiple stages of the study, resulting in different types of data. The results are organised around the study's research questions. Portfolio assessment data are first presented to examine the extent to which students developed and demonstrated TVCs during the programme. Particular attention is given to the quality of assessment evidence and students' justifications, as these provide insight into the validity of the assessment data used in competency judgements. The focus group results are then presented to explore participants' perceptions of the assessment approach and to identify learning processes that supported competency development. Finally, the findings are synthesised to identify prerequisites and critical success factors for implementing programmatic assessment of transversal competencies in aircraft maintenance training.

ISSN 3006-6301, Vol. 3 (No. 2) 2026

Student Portfolios

Every week, students completed one SA and PA on a specific OB. Instructors were responsible for monitoring all 21 students, typically dividing them into groups and rotating responsibilities weekly. However, instructors were unable to collect comprehensive assessment data for all students within a single lesson. Consequently, there was a greater volume of peer- and self-assessment data for each student. Sometimes there were duplications in PAs as students failed to make clear arrangements about who assessed whom, resulting on one student having two peer assessments and another student missing one. In this case, for the student with two PAs, the assessment with the most valid examples was included in the portfolio.

Portfolio Quality

Beyond examining the ratings themselves, the analysis focused on students' justifications, as these provide insight into how students interpreted observed behaviours and whether they were able to connect these observations to the appropriate rubric criteria and performance levels.

The quality of the examples the students provided to support their ratings varied across the students. For both PAs and SAs, four types of justification patterns were identified:

1. No justification was provided
2. A weak justification was provided
3. A valid justification was provided
4. The justification did not align with the rating, most often because the student was overly strict with themselves or with a peer.

However, overall, the students improved in providing valid examples to support their ratings for both the SAs and PAs. At the start, they tend to be very generic or just repeated the question through a statement. Later, their examples were more detailed and referred to specific moments during their practical work in the hangar. Sometimes students did not provide any examples to support their ratings. Table 4 presents examples of students' justifications for rating their own behaviour as 'competent' for selected rubric elements. Table 5 provides examples for PA ratings of 'competent,' using different rubric elements to offer a broader perspective.

Table 4: Examples of poor and strong justifications for competent ratings on self-assessments

Observable Behaviour	Description of Competent Behaviour in Self-Assessment	Examples poor justification	Examples strong justification
11.4	I communicated only essential information that was important for safety. I did not communicate unnecessary information.	*"I only communicate essential information when we talk about work."*	*"I communicated to my team that we needed to put the thrust reverser on safe without any irrelevant additions."*
8.1	I discussed all mistakes or errors I made and involved all team members, including the instructor/supervisor.	*"I don't make mistakes."*	*"After I overtightened a bolt, I promptly discussed the issue with my peers, the instructor and the supervisor to jointly find a solution."*
8.12	I usually or always responded to non-verbal signals, such as body language from fellow students or colleagues, indicating they needed help.	*"I helped him with drilling."*	*"When we were jacking up the aircraft, I noticed my peer student got tired and I offered to switch and take over."*

Table 5: Examples of poor and strong justifications for competent ratings on peer assessments

Observable Behaviour	Description of Competent Behaviour in Peer Assessment	Examples poor justification	Examples strong justification
11.4	Your peer student checked whether they understood you correctly.	*"Yes, he did check to see if he understood me."*	*"When he was in the cockpit and I was on the floor he double-checked everything I said to be sure what I meant, and repeated the steps back to me for confirmation."*
8.1	Your peer student involved all relevant colleagues when discussing something.	*"Yes, he told me what we were going to do."*	*"Yes. During the hydraulic testing he involved all students around the aircraft, so they knew exactly what was going to happen."*
8.12	Your peer student offered their help proactively	*"He always helps me."*	*"When I struggled understanding the manual, he asked if I needed help and explained what the manual said."*

Instructors always provided concrete examples to justify their ratings, based on their notes taken during classes.

ISSN 3006-6301, Vol. 3 (No. 2) 2026

Rating as Beginner or In Development

While the previous section discussed students' justifications for competent behaviour, many ratings fell below that level. The most reflective justifications appeared in SAs rated as 'In Development'. Interestingly, comparable justifications were sometimes associated with different self-ratings: students who rated themselves as *'In Development'* occasionally provided reflections similar to those offered by students who rated themselves as *'Competent'*. Table 6 provides examples of reflective justifications for ratings below *'Competent'*.

Table 6: Examples of justifications for students rating themselves as 'In Development'

Observable Behaviour	Description of 'In Development' Behaviour in Self-Assessment	Justification of rating 'In Development'
11.4	I sometimes assumed that the information I communicated was clear to the recipients without checking, or I checked only with a few recipients.	*"I assumed my team members understood their tasks in the cockpit, but I didn't check this with control questions, so I couldn't be sure."*
8.1	I sometimes discussed mistakes or errors with my fellow students/colleagues or instructor/supervisor or limited this to one or a few team members.	*"There had been confusion during the greasing process about what steps we had taken. We should have communicated this better from the start."*
8.12	I sometimes responded when my fellow student/colleague needed help while spending a very long time on a task.	*"Sometimes I was busy myself so I couldn't really help. When I noticed my colleagues took long completing their tasks I tried to help them. But there is a pitfall; because I was doing something else myself, I wasn't always sure what my peer students were doing."*

Holistic Portfolio Evaluation

To ensure an accurate assessment of each student's competence, it was necessary to have sufficient data for every observable behaviour. Consequently, two students were not included in the portfolio evaluations: one was absent frequently owing to personal circumstances, while the other missed numerous classes because of motivational challenges. For the remaining nineteen students, final assessment was determined based on both the scores obtained during the assessment weeks and any notable observations made during or after this period. Such observations could be positive, for example, when a student demonstrated competence in a learning objective at a later stage, or negative, such as when earlier scores appeared to be adversely influenced by inappropriate behaviour or by clearly missed learning opportunities. Additionally, each rating must be substantiated by appropriate, valid examples. Ratings without support of clear examples were excluded from the dataset. The overarching aim of this approach was to construct a reliable and holistic profile of each student, enabling the committee to make a reasoned estimation of future performance and professional behaviour.

The degree and quality of student reflection were also considered. Students who assessed themselves as 'Beginner' or 'In Development' and provided a clear, critical description of areas for improvement were considered to demonstrate at least as much, if not more, competence than students who rated themselves or their peers as 'Competent' with limited justification. Several students appeared particularly self-critical due to their reflective disposition. Although they were already performing well in many respects, their reflections focused predominantly on areas requiring further development. The substantiation they provided, including evidence of what they were doing effectively, was often more robust than the justification offered by other students for assigning themselves a 'Competent' rating. On this basis, the committee adjusted these comparatively low self-assessments to a 'Competent' level, recognising that these students had applied stricter evaluative standards to their own performance.

Based on these considerations, in total three students were considered competent on all three OBs after nine weeks of assessments. Adding the extra assessment opportunity given, then in total five students were considered competent on all OBs. Other students were considered competent on one or two OBs, and some needed more time to reach that level. Table 7 shows the final results regarding the portfolio assessments, where the extra assessment option has been included.

Table 7: Portfolio Assessment Results

Observable Behaviour	**Number of students deemed competent**
8.1 *Fosters an atmosphere of open communication*	11
8.12 *Anticipates and responds appropriately to the needs of others'*	15
11.4 Maintains situational awareness when selecting method of communication, speaks clearly, accurately and concisely.	6

Focus Group Results

For the focus group, first the perception of various assessments is evaluated, followed by the learning effects of the training and assessment programme. Finally, the prerequisites identified by both students and instructors as essential for successful implementation are discussed. The students are anonymous and are referred to by a number that corresponds to their place in the seven different groups.

Perception of the Self- and Peer-Assessments

All students confirmed that completing the assessments properly is something that can be expected from EQF level four students. However, they explained that it does take effort for mainly two reasons. The first reason is reading, understanding and interpreting the text. Student 5 explained: *"I think some students did not always understand what was written in the rubrics, and if you don't really understand well then it's difficult to fill it in"* although later, when the examples of some justifications were discussed, this student added *"it can also be laziness, if you put just some effort into reading the text properly…."*. The other reason the students experienced difficulty, was reporting valid and concrete examples to justify their ratings. Student 2 mentioned: *"Sometimes writing down the examples came almost automatically, while other times it was really difficult to think of a concrete example of the*

ISSN 3006-6301, Vol. 3 (No. 2) 2026

behaviours". Sometimes the effort taken into reading the assessment assisted in finding examples, as student 17 explained:

> Sometimes I was reading it and then I just had to read it a few times to make sure I understood. And then while I was reading it, a concrete example from the lesson just came to mind and then I could write that down.

Some students extended their reflections to outside the classroom. As student 1 explained:

> I know the assessments reflected on the lesson of that day, but I also looked at my daily life. You simply cannot rate yourself as 'Competent' if you only do it that day and then when you come home you stop doing it. So, I also thought about how I was at my work and my internship. And then if I did it there as well, I rated myself as competent. If not as 'In Development' or 'Beginner' even.

Students often based their ratings on their described examples, instead of the other way around. Students 17, 10 and 13 explained: (17*) "I did not always rate first and then wrote down the example. Sometimes I did it the other way around"* (10*) "I always did it the other way around.* (13) *"I had the same. First, I wrote down the example and then I gave the rating I thought was best".*

During the session, several anonymized assessment examples were reviewed, comparing justifications for the ratings 'Competent' and 'In Development'. This process highlighted inconsistencies in the interpretation of evaluation criteria. The students collectively observed that those who articulated their areas for improvement demonstrated a higher degree of competence than the students who rated themselves as competent without providing substantial justification.

When asked about anonymity, the students indicated that they were not worried about the fact that the peer assessments were not anonymous. As student 5 explained: *"I think it's not unusual, especially in aircraft maintenance, to have someone watching or, yes, in some form, offering criticism or improvements to your work and give feedback. So, I didn't find it particularly strange"*

Objectivity

Several students reported challenges in maintaining complete objectivity during peer-assessments, especially when evaluating peers who are also friends. Nevertheless, these students explained that withholding honest feedback would ultimately not benefit their friends, emphasizing the importance of providing constructive opportunities for growth. As student 17 mentioned:

> You are in the same class for several years already, so then it's sometimes difficult to be really critical. I did it though, also towards my friends. And I told him when he did something wrong. And then maybe he will do it better the week after.

Student 10 added:

> At some point student 11 just didn't score above 'Beginner' so I just had to rate him as such. But at the same time, you think "he is so close to being 'In Development'". That was tough, especially when you are friends, of course.

Students 1 and 5 explained: (1*) "When in doubt I tended to give lower. If I rate too high, then it's not good for him."* (5) "*Yes, if it's only roses and moonlight and you give the other person the feeling that all is well while that's not the case, well, then, in the real world he will be held accountable for it."*

Other students reported no issues with objectivity, as student 1 stated: *"I really didn't mind, I just followed exactly what the rubric said and rated accordingly."*

Confidence

When students were asked to what extent they believed their judgments were accurate, they demonstrated considerable confidence. The more time students spent on a certain OB, the greater their confidence became in understanding and assessing their own performance and that of others. As student 13 explained:

> Reflecting on the same topic week after week made it easier to understand what I was doing and to see my progress. If I wasn't sure what to focus on, returning to it regularly helped me improve. When we started reflecting on a new topic, though, I had to shift my perspective, which took much more time. Each subject required a different approach, but reflecting in this way was valuable, not just for learning about myself, but also for gaining insights into those around me.

Many students echoed the same sentiment in different words, and all students agreed that they would have preferred more time for each OB. As student 5 explained:

> Sometimes, I was just getting the hang of it and then I had to move on to the next assessment. I would have preferred one more lesson where I could have practiced it a bit more, and then... I just always thought, "Well, now I've got it down pat."

All students confirmed that discussing the validity of certain examples during the theory lessons really helped them to build confidence over the quality of their examples and ratings.

When asked about the decision making when filling in the assessments, the students also demonstrated confidence although they were honest about their doubts during the decision-making process for PAs. Student 15 explained:

> You see two sides of your peer. You see somebody who works seriously and then you think: "well he does everything that's expected" but then he shows another side where he is pushing the boundaries and then I found it difficult to make a judgement.

Student 5 added:

> Sometimes I found it hard to judge the boys' performance because sometimes they did what was expected, but often they didn't. They really didn't. And then he did show something, but was I convinced? This made it difficult to decide how to assess them. I wasn't sure whether to criticise them fully since they did make some effort, but not enough. I struggled with what to write. Eventually, what was on paper, yes, I was very certain of that. But before I put it down on paper, I had to weigh certain things before reaching a sound judgement that I stand by.

Motivation

Students were asked whether they felt motivated to score competent themselves. In their responses, the students demonstrated a strong focus on growth rather than passing. Therefore, they were honest in their ratings. As student 5 explained:

> Passing is nice, but it's not why you do it. If you are really critical towards yourself, then you admit that there are still things to learn. That is much more valuable than passing. So, I am not going to rate myself 'Competent' thinking "then I pass" while my actual performance was insufficient. Then you're cheating.

Student 13 added: "*It's not that you want to rate yourself 'Competent', it's more that you want to improve yourself and you put your efforts in that. And you want to involve your environment in that."*

When asked whether the fact they were being assessed affected their behaviour during classes, the students answered negative. Student 15 explained:

> I think, if you start acting, then you're just fooling yourself. If you only do it when the instructor stands next to you and you think "now I need to do this" and the rest of the time you don't, then you are just fooling yourself, pretending you're better than you are.

Student 10 added: *"You can't keep that up for 3 weeks anyway".*

Perception of the Instructors

Both instructors acknowledged the complexity of assessing a group of students on behaviour and performance. As one instructor explained: *"If you could just check-off a list of behaviours then it would be easy but that's not what you want. You really want to describe the behaviour and then analyse the rating. I found that difficult."* The other instructor agreed that the ORCA method is a prerequisite to really evaluate behaviour:

> If you do it the other way around, you become lazy. You run the risk of the "no news is good news" attitude. But that's not enough. If somebody performed well, you want to describe why it is considered well.

This instructor also explained the dependence on the situation when something can be evaluated: "*We're in a hangar with 21 students so of course you have blind spots. But other times you catch a problem and then you can easily assess the student's performance. It really depends on the situation."* He added how the role-play supported the elicitation of the behaviours: "*putting students in a situation where I can test whether they just assume something or not was very easy to do".* How students experienced the role-play elements will be explained in the next section.

Learning Effectiveness

The students shared many insights about what they gained from the training and assessment programme, focusing mostly on reflection, increased awareness, and how these factors influenced their performance going forward. Table 8 represents the categories and the examples given by students regarding the learning impact, voiced in different ways and / or confirmed by their peers during the session.

Not only did the students explain what they learned from the programme, they also mentioned which aspects of the programme triggered and / or supported their competency development, covering the following categories:

- Role-play and simulations
- Instructors as coaches and assessors
- Iterative process
- Collaborative learning
- Formative feedback rounds
- Other didactics

Table 8: Summary of outcomes and effectiveness of the training and assessment programme

Reflection and Awareness	Future Performance
"Reflecting on myself and others and put that into words and explaining someone else what he didn't do well."	*"Later as Certifying Staff I will also have people working under me and now I really learned to look at my own behaviour but also how a team functions."*
"I am now more aware of my approach during collaboration."	*"I will be more mindful of how I interact with others and allow them to discover things independently."*
"Getting to know myself better, how I respond to and deal with certain situations I never encountered before."	*"Because I have been brought into certain situations already, I can now anticipate my behaviour."*
"Understanding what I was doing and monitoring my progress."	*"During my internship, I realized that some of the competencies I'm learning now are situations I've encountered before. Reflecting on those moments, I see how I'd respond differently today."*
"Reflecting on whether my communication was clear and effective, and vice versa."	*"I can adjust my language towards others, reducing the chance of misunderstandings."*
"By consciously reflecting on them I realized I make way more assumptions than I thought."	*"I will recognize earlier now that I am making assumptions."*
"Relating theoretical knowledge about the Dirty Dozen to what was happening in practice."*	*"For example, I've learned to stay neutral and provide clear, unbiased answers, even when asked obvious questions. This approach supports open and impartial communication within a team or company. That's something I want to continue doing."*

Note: the Dirty Dozen are twelve factors associated with maintenance related aviation accidents and incidents and therefore addressed in the theoretical module "Human Factors": Lack of Communication, Complacency, Lack of Knowledge, Distraction, Lack of Teamwork, Fatigue, Lack of Resources, Pressure, Lack of Assertiveness, Stress, Lack of Awareness, and Norms (Transport Canada, 2003).

Roleplay and Simulations

All students confirmed that including role-play in the practical lessons greatly enhanced their competency development. As student 5 illustrated:

> I found it especially educational. I didn't always enjoy it, in fact, sometimes it was just annoying, and it felt like someone was constantly on your case. But it was genuinely very educational, because I was put into situations I never really experienced before. During previous work placements, maybe a little, but I wasn't actually responsible, so it wasn't really my problem, and I didn't worry about it. But this time, it did become my problem, and I had a team behind me, or someone I had to deal with, which made it all the more instructive.

Student 15 added:

> The instructor, for example, played an intern who knew nothing. Yes, that's something that can really come back later in your work. But in a normal practical lesson, you're all reasonably on the same level, so suddenly having someone who actually knew nothing and dropped things…. how do you deal with that? And how do you respond to that? I find that very interesting.

Instructors as Coaches and Assessors

The students were all very much in favour of how the instructors approached them and challenged them. As illustrated by student 15:

> In the past few months, I learned more than in the three years before. Not just because of the competencies but because of the whole setting. Of course, we already had some theoretical knowledge, but when I asked something you all really helped me and coached me in a certain direction by asking questions that would lead me to the right answer. You didn't just refer me back to the books. That was really effective.

Furthermore, having the instructors' eyes on them while in a simulated setting was experienced as stimulating, as explained by student 5:

> It was helpful that in the practical lessons, there was sometimes someone right behind you, watching you, or pushing you into a situation you wouldn't normally encounter in such a practical lesson. I find that instructive because you see a side of yourself that normally remains underexposed in practice, I think.

Student 13 added:

> With the instructors next to you, you feel under pressure. That pressure highlights weaknesses, which are good opportunities to improve. For example, with Instructor 1, I was tested on the solenoid but realized I didn't know the answer. Although stressful, such moments are valuable for making progress.

Iterative Process

As previously stated, every student stressed how vital the iterative nature of programmatic assessment is for improving their performance, building confidence, and fully grasping the standards needed to be deemed competent. In addition to what has been quoted earlier, students 2 and 10 explained: *"I found*

it very advantageous that each topic is assessed for a few weeks. This also gives us the opportunity to improve ourselves. Yes, to critique. So, we really experience the topic in practice." And: "*Although there sometimes wasn't much feedback from instructors, making the same assessments each time and seeing the peer assessments helped me learn and improve independently."*

Formative Feedback Rounds

All students confirmed that the feedback rounds, where the assessments were discussed in small groups with three students and the instructors, were very powerful. Student 13 worded it as follows:

> Those group discussions were very useful for us, because then you could really connect with each other and come together as a team, but also discuss things individually. It strengthens team bonds and helps everyone to stand stronger. That's really good for that.

In addition, some groups also discussed feedback outside of the feedback sessions with instructors, as indicated by student 10:

> We discussed it every time. That went well. I did this with student 11 and student 12, and you could really see the progress being made; they were learning from it. I discussed things with them, and when I explained, student 11 said, "Oh yes, I see it now." And then he paid more attention the next time, which meant I could give him better feedback afterwards. I appreciated that. I enjoyed doing that.

Overall, however, students reported that they would not suggest leaving feedback discussions solely to students themselves, preferring instead to have these conversations with instructors present. Their main reasons were concerns about social safety and consistency; without instructor oversight and organization, some students might avoid participating in the discussion altogether.

Other Didactical Benefits

Other approaches that students mentioned as supporting competency development were the evaluation of the assessment examples during theory lessons, rotate leadership within their team, practicing the TVCs while performing an A-check on the aircraft, and the theoretical lessons. As explained by student 13:

> The theoretical lessons really served as a foundation before going into the hangar. They explained the basic principles of what we were going to do. If you don't have that knowledge, then you don't know what you are doing. For example, when we talked about the Swiss Cheese Model. You really had to understand that when mistakes were made in the hangar. And then we discussed that; mistakes made by people. If you don't know that at the start, then you can't do the assessments.

Student 2 added: *"The theory helped me describe and focus on what we had to do."* The students found the rubric-videos particularly useful to understand what good and bad performance looks like; *"For me that helps better than reading it in a book, because theory and practice, it's just not the same. So, this way I could remember it better"* student 13 clarified.

Prerequisites for Successful Implementation

When asked what the students considered pivotal for successful implementation, the following elements were stressed:

- Role-play and simulations
- Coaching approach of instructors
- Repetitive nature of the assessments
- Formative feedback sessions with instructors

In addition, students suggested several changes to improve the effectiveness of the programme. They indicated that reflection on performance through assessment should begin as soon as students enter the college, so that all students become familiar with the process and learn to take it seriously from the outset. They also recommended adjusting group composition as students progress, allowing more advanced students to support those who need further development. This approach was implemented during the final optional assessment week and was perceived as valuable, as repeatedly working in the same trio was seen to limit opportunities for growth. Furthermore, students emphasised the need to allow more time for each OB to be assessed repeatedly. Finally, they suggested that, as the programme advances, more targeted assessments should be conducted to address specific competencies, with groups formed to evaluate areas where additional data is required, focusing exclusively on those elements that need further analysis.

Notably, most students did not view the compulsory integration of TVCs for licensing as essential to their motivation for achieving a competent score. When questioned about whether mandatory requirements for the Aircraft Maintenance License would have influenced their effort, student 10 responded: *"Maybe I would have concentrated more effectively during the initial week. Because I noticed at the start, we were a bit laxer about it."* Most students echoed student 1 who said: *"I would have done exactly the same as I did now."* Student 15 added a different statement: *"you know it's just a pilot. So, I think we would try harder if it was real".*

The instructors emphasized the importance of assessor training. As instructor 2 outlined:

> If you have to do this full-time as an instructor, you must acquire certain qualities that I sometimes didn't even possess myself, to accurately describe what I'd seen. You wonder "Is this something I'm actually noticing now, or am I just assuming it happened because of the current situation?" And then writing that down clearly. Yes, I think as instructors we will all do this, so to speak.

Discussion

This section provides a more in-depth discussion and analysis of the results. The two stages of data collection remain in structure but will be more closely linked to gain deeper understanding of their implications.

Student Portfolio

Portfolio quality varied among students. While some provided detailed examples to support their ratings, others were less advanced in doing so. Students who gave thorough peer evaluations could receive weakly supported feedback in return, which influenced their overall portfolio assessment data reliability. Students attributed this issue to both a limited understanding of the texts and a lack of motivation to invest effort. Although most students stated that not being required for their licence didn't impact their approach, this factor may still have reduced the motivation for several other students.

As mentioned in the results section, reflective justifications came from students rating themselves as 'In Development', while some other students rated themselves as 'Competent' substantiated with a lower performance level description. This suggests inconsistency in how students interpret the threshold descriptions. As highlighted in the introduction, students must develop familiarity with assessment terminology, learn to provide feedback, and accurately and consistently interpret assessment criteria (Kearney, 2013; Lerchenfeldt et al., 2019; Resta & Lee, 2010). Therefore, this process should begin early in their training (Tornwall, 2018). During the Focus Group session, the students highlighted the same message, and the portfolio data underscored that some students need more time to effectively benefit from PAs and SAs to evaluate TVCs. Nevertheless, although TVC assessments were newly implemented for these students, data indicated that over the nine-week period, students already demonstrated progress in offering valid justifications for their ratings. Additionally, students reported an increased understanding of the competencies as the weeks progressed, improving their competency performance.

In some cases, students had not coordinated the peer assessments properly, resulting in duplications and omissions of assessment data. A pre-established assessment schedule could address this issue; however, in practice, unforeseen student absences would have limited its effectiveness.

Peer-Assessment Preferences in Aircraft Maintenance Training

Although the discussed literature states that most students prefer PAs to be strictly anonymous, the aircraft maintenance students reported no issue with the fact that this wasn't possible. Moreover, students reported a preference for regularly changing group composition to ensure an optimal mix of competency levels and to further support competency development. This contrasts with the literature, which underscores the importance of teams working in stable groups for several months.

Competency Development

The students reported a high degree of learning effects resulting from the training and assessment programme, mainly focusing on reflection, awareness and understanding what behaviour is required regarding communication and teamwork. Furthermore, as a result, they reported how it affects their future performance as maintenance engineers. This aligns with the literature discussed in the introduction, stating that PAs and SAs contribute to communication, teamwork, metacognition, reflection, awareness, and an understanding of standards and criteria (Kearney, 2013; Lerchenfeldt et al., 2019; To & Panadero, 2019). Additionally, peer- and self-assessments foster a more open and supportive class climate (Ndoye, 2017). This is of particular interest for the OBs of teamwork, addressing open communication and responding to the needs of others. Not only are these OBs extensively addressed and monitored, but the method itself may also have contributed to the development of these competencies. This could potentially explain why the portfolio evaluation showed that, following the training and assessment programme, a substantially higher proportion of

students demonstrated competence in the teamwork OBs compared to the communication OB. Another possible explanation relates to sequencing: students began with OB 11.4, which was the first TVC introduced in their technical training and the first instance in which they were required to complete PAs and SAs. Finally, OB 11.4 is considered the most crucial TVC of ICAO's competency framework and the most challenging to assess (Kes, Saunders-Smits, et al., 2026). This may explain why students need more time to develop towards the required performance level for this OB.

Students placed greater value on growth and development than on achieving high ratings. This finding aligns with previous research involving focus group sessions with industry stakeholders, which suggested that successful implementation of CBTA requires a shift from a pass/fail orientation toward a growth mindset (Kes, Pietersma, et al., 2026). The results of the present study may indicate that participating students are already adopting this developmental perspective, whereas parts of the more established maintenance workforce in the aviation industry may still be in the process of making this transition.

Limitations

The average student portfolio contained three PAs and three SAs for each OB, supplemented with instructor assessment data although in a smaller volume. While most students showed improvement providing valid examples to support their ratings, since the assessments concerned a four to six hours period in the hangar, it was not always the same behaviour evaluated which resulted in a variety of ratings and examples per day. The most consistent portfolios are those where the ratings are consistent but with a variety of supporting examples to underscore the performance levels. Considering that this is the first time for students being assessed on TVCs and using PAs and SAs, the daily performance of many students largely fluctuated, making it more difficult to predict future performance. This suggests that greater consistency in the performance context may be needed to facilitate more reliable tracking of competency development over time. Students were assessed during different tasks and interactions, meaning that observed behaviours were not always directly comparable across assessment moments. While a more standardized context could improve consistency, it may also reduce authenticity by creating more staged and less natural workplace situations.

A further limitation concerns the sequencing of training the OBs. Although students' performance before and after the focal OB was considered in the final portfolio assessment, competency development does not progress in isolation. As training advances, OBs increasingly interact and mutually reinforce one another. This interdependence was already evident in the way the assessment approach itself appeared to facilitate growth in OBs 8.1 and 8.12. A longer and more comprehensive study would enable a more holistic examination of how these OBs develop in parallel and influence each other over time.

Also, this study represents an initial step in implementing TVCs within a holistic technical training environment, aimed at exploring both their perceived feasibility and effectiveness. The findings are based on a single cohort of students at a public college in the Netherlands, which limits the generalisability of the results. To gain a more comprehensive understanding of how TVCs can be effectively trained and assessed within a CBTA context for aircraft maintenance, similar programmes should be implemented and evaluated across diverse national contexts, training organisations, and educational institutions. Nevertheless, as stated earlier, Cain (2011) indicated that context-specific findings from AR can still inform and inspire similar practices in comparable settings.

Finally, the dual role of the first author in both programme implementation and data collection and analysis may have influenced the interpretation of the findings. However, such close involvement is

inherent to AR, where researchers actively participate in the intervention process rather than remaining detached observers. While this embedded position enables a deeper, contextually grounded understanding of the educational practice under study, it also requires careful reflexivity when interpreting the results.

Conclusion

This study aimed to evaluate the perception and effectiveness of the developed and validated training and assessment programme, which integrates the TVCs communication and teamwork within the technical context of aircraft maintenance. First, each of the sub research questions will be addressed, which collectively inform the answer to the main research question: ***What are critical success factors for training and assessing transversal competencies in aircraft maintenance VET using programmatic assessment principles?***

Sub-question 1: Perception

Sub-question 1 addresses the perception of SAs, PAs, and IAs within the programmatic assessment approach. Students from EQF level four confirm that the SAs and PAs correspond to their training level, but that they had to put effort into both understanding and completing the assessments properly. Less motivated students may therefore produce less valid assessment data both for their peers and themselves. Although students acknowledge objectivity challenges when assessing friends, their focus on the importance of growth rather than passing a test, results in honest ratings and constructive feedback. Students feel confident in their ability to complete the assessments. They note that the process occasionally requires substantial weighing of evidence, but they remain confident in the final scores they award and in their accompanying justification.

Instructors report that assessing many students is challenging, especially since competency assessment should be done holistically, rather than checking-off a list of behaviours. Although conducting the A-check provides sufficient opportunities to assess the TVCs, recording and evaluating the performance remains complex.

In summary, the training and assessment programme and the programmatic assessment approach is perceived as feasible for EQF level 4 students. However, instructors need further support in overseeing larger groups of students while evaluating their performance holistically.

Sub-question 2: Training Effectiveness and Competency Development

Sub-question 2 addresses to what extent the training and assessment programme enhance and support the development of transversal competencies. Both the student portfolios and the focus group results demonstrate an increasing understanding of what performance is required as time progresses. The students report an increasing awareness and ability to reflect on their own communication and teamwork competencies and those of their peers, corresponding with their portfolio data which showed that over time, students improved in providing valid, specific examples to support their self- and peer-assessments, moving from generic or repetitive statements to more detailed reflections grounded in concrete experiences during practical work in the hangar.

Additionally, the students affirm how the training and assessment programme positively affects their future performance as maintenance engineers. The assessment of the teamwork competency shows more progress and higher results than that of communication. This may be partly due to the method itself, which is known to enhance teamwork and open communication. However, other factors, such as the sequence of the training or the inherent complexity of communication, may also explain why progress in communication lags behind improvements in teamwork.

Sub-question 3: Key Prerequisites for Successful Implementation

Sub-question 3 explores the key prerequisites for effectively implementing programmatic assessment of TVCs in competency-based aircraft maintenance training. The following pedagogical, didactical and organisational aspects are considered key prerequisites both by students and instructors.

Early exposure to TVC assessments: Students should be introduced to reflective performance assessment from the moment they enter the college. Early exposure allows them to develop familiarity with assessment expectations, understand the seriousness of reflective practice, and cultivate habits of self-- and peer-evaluation that will support their long-term growth.

Role-play and simulations: Another foundational element is the consistent incorporation of role-play and simulation within authentic technical contexts. These activities create realistic scenarios in which students can demonstrate observable behaviours and receive feedback on their communication, teamwork, and technical performance.

Coaching instruction approach: Instructors should adopt a coaching-oriented approach. Rather than simply correcting errors, they guide students in analysing their performance and identifying areas for development.

Iterative process: To support continuous learning, the repetitive nature of programmatic assessments must be maintained. Each OB should be assessed multiple times to capture progression over time. Adequate time must be allocated for students to engage with each OB as a new developmental journey.

Formative feedback sessions: Formative feedback is a critical component of this process. Feedback sessions should take place in small, socially safe groups, facilitated by instructors. Such settings help students share experiences, discuss peer-assessments, and receive constructive guidance without fear of judgement.

Variety in group composition: As students advance through the programme, group composition should be adjusted deliberately. Mixing students with varying competency levels enables more experienced learners to model effective behaviours while providing those needing improvement with new learning opportunities. Keeping the same groupings for too long, especially after extended collaboration, may hinder growth.

Targeted assessments: In later stages of the programme, assessments should become more targeted. Groups may be formed with the explicit purpose of evaluating specific competencies or collecting additional assessment data, allowing instructors to focus on elements that require further analysis.

Instructor requirements: To ensure quality and consistency in assessment, instructors and assessors should apply the ORCA method. This structured approach supports objective observation, disciplined documentation, and well-substantiated justification of ratings, thereby improving the validity of the assessments and the quality of feedback.

Training requirements: Finally, sufficient training must be provided to instructors and assessors. Holistic performance assessment in larger groups demands calibrated judgement and familiarity with behavioural indicators.

The preceding list, as found in this study, is a first indication of critical success factors for training and assessing TVCs in aircraft maintenance VET. Establishing these factors represents an important step towards implementing CBTA in aircraft maintenance training, particularly when TVCs become integral components of the Aircraft Maintenance License. However, further research is needed to determine the extent to which the developed training and assessment programme can be transferred to industry training organisations and other education and training institutes in different countries. Additionally, the potential for interplay between OBs, where progress in one OB may naturally stimulate development in others, warrants deeper investigation. Understanding these interactions could clarify whether certain OBs can be strengthened indirectly, without requiring separate, targeted instruction.

Acknowledgements

This study was supported by MBO College Airport, VET college for Airport & Aviation, part of ROC van Amsterdam – Flevoland. The authors would like to thank Ben Pietersma, Rene de Moor, Gerrit Kunst and Yara van Oosterhout for their active contribution to the study. We also gratefully acknowledge the students who participated in the research.

Declaration of Interest

There are no conflicts of interest to disclose in connection with this research.

Notes on Contributors

Lydia Kes is a PhD candidate at Delft University of Technology, focusing on transversal competency assessment in aircraft maintenance training. She is Head of Research & Development at MBO College Airport and an adviser to ICAO's Personnel Training and Licensing Panel. She holds a master's degree in Neuropsychology.

Marcus Specht is Full Professor of Digital Education at Delft University of Technology and Professor of Learning Sciences at FernUniversität Hagen. His research focuses on digital learning and educational technologies. Previously, he led the Mobile Knowledge department at Fraunhofer FIT and was Professor of Learning Technologies at Open Universiteit Nederland.

David Abbink is Full Professor of Haptic Human–Robot Interaction at Delft University of Technology. His work focuses on shaping the future of work with robotics through transdisciplinary research. He leads large-scale initiatives on human–robot collaboration and received the NWO Stevin Prize (2024) for his contributions.

Gillian Saunders–Smits is Associate Professor at the Cognitive Robotics Department, Faculty of Mechanical Engineering, Delft University of Technology. She holds an MSc and PhD in Aerospace Engineering. Her research focuses on engineering education, including curriculum development, innovative teaching, labour market preparation, inclusive education, and online learning.

References

Al Mamun, A. (2012). The Soft Skills Education for the Vocational Graduate: Value as Work Readiness Skills. *British Journal of Education, Society & Behavioural Science*, *2*(4), 326–338.

Amin, N., Yother, T. L., Johnson, M. E., & Rayz, J. (2022). Exploration of Natural Language Processing (NLP) Applications in Aviation. *Collegiate Aviation Review International*, *40*(1). https://doi.org/10.22488/okstate.22.100211

Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2006). The wheel of competency assessment: Presenting quality criteria for competency assessment programs. *Studies in Educational Evaluation*, *32*(2), 153–170. https://doi.org/10.1016/j.stueduc.2006.04.006

Biggs, J. B. (1996). Enhancing teaching through constructive alignment. *Higher Education*, *32*, 347–364.

Boeing. (2022). *Boeing Pilot and Technician Outlook 2022—2041*.
Cain, T. (2011). Teachers' classroom-based action research. *International Journal of Research & Method in Education*, *34*(1), 3–16. https://doi.org/10.1080/1743727X.2011.552307

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (Eighth edition). Routledge.

Costanza, D., Prentice, B., Poitras, M., Bardygula, O., Fasano, B., Sargent, S., Hayes, L., & Franzoni, C. (2022). *Not Enough Aviation Mechanics. How the industry can address this decade's shortage in aircraft maintenance workers.* Oliver Wyman. https://www.oliverwyman.com/content/dam/oliver-wyman/v2/publications/2023/jan/Not%20Enough%20Aviation%20Mechanics%20-%20AMT%20report.pdf

EASA. (2022). *Easy Access Rules for Continuing Airworthiness (Regulation (EU) No 1321/2014)*. European Union Aviation Safety Agency. https://www.easa.europa.eu/en/document-library/easy-access-rules/easy-access-rules-continuing-airworthiness-0

Farland, M. Z., & Beck, D. E. (2019). Collaborative Learning Teams to Longitudinally Teach and Assess Teamwork Behaviors and Attitudes. *American Journal of Pharmaceutical Education*, *83*(9), 7255. https://doi.org/10.5688/ajpe7255

Gerritsen-van Leeuwenkamp, K. J., Joosten-ten Brinke, D., & Kester, L. (2019). Students' perceptions of assessment quality related to their learning approaches and learning outcomes. *Studies in Educational Evaluation*, *63*, 72–82. https://doi.org/10.1016/j.stueduc.2019.07.005
Gibson, M., & Arnott, D. (2007). The Use of Focus Groups in Design Science Research. *ACIS 2007 Proceedings 14*, 327–337. http://aisel.aisnet.org/acis2007/14

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2007). Defining authentic assessment. Five dimensions of authenticity. In *Balancing Dilemmas in Assessment and Learning in Contemporary Education.* (pp. 75–88). Routledge. https://doi.org/10.4324/9780203942185

Gulikers, J. T. M., Runhaar, P., & Mulder, M. (2018). An assessment innovation as flywheel for changing teaching and learning. *Journal of Vocational Education & Training*, *70*(2), 212–231. https://doi.org/10.1080/13636820.2017.1394353

ICAO. (2020). *ICAO Doc 9868. Procedures for Air Navigation Services. Training*.
Ichou, S., & Veress, A. (2023). Maintenance 4.0: Automation of Aircraft Maintenance Operational Processes. *International Journal of Aviation Science and Technology*, *vm04*(is01), 23–31. https://doi.org/10.23890/IJAST.vm04is01.0103

International Air Transport Association. (2023). *Competency Assessment and Evaluation for Pilots, Instructors and Evaluators—Guidance Material*. IATA.

Jones, D. K., & Abdallah, M. (2013). *Assessment of Communication and Teamwork Skills in Engineering Technology Programs*. 23–225.

Kearney, S. (2013). Improving engagement: The use of 'Authentic self-and peer-assessment for learning' to enhance the student learning experience. *Assessment & Evaluation in Higher Education*, *38*(7), 875–891. https://doi.org/10.1080/02602938.2012.751963

Kearns, S. K., Mavin, T. J., & Hodge, S. (2016). *Competency-based education in aviation: Exploring alternate training pathways*. Ashgate.

Kemmis, S. (2009). Action research as a practice-based practice. *Educational Action Research*, *17*(3), 463–474. https://doi.org/10.1080/09650790903093284

Kes, L., Pietersma, B. S., Saunders -Smits, G. N., & Specht, M. M. (2025). Valid Assessment of Transversal Competencies in Aircraft Maintenance: Insights From Focus Groups With Industry Stakeholders. *ATEC Journal*, *47*(2), 26–45.

Kes, L., Pietersma, B., Saunders-Smits, G. N., & Specht, M. (2026). *EXPLORING COMPETENCY-BASED TRAINING AND ASSESSMENT: INCLUDING TRANSVERSAL COMPETENCIES IN VOCATIONAL AIRCRAFT MAINTENANCE EDUCATION*. 1097. https://doi.org/10.21125/inted.2026.1097

Kes, L., Saunders-Smits, G. N., & Specht, M. M. (2026). Towards new educational standards for aircraft maintenance: Analysing transversal competencies on industry safety priorities and assessment challenges. *Cogent Education*, *13*(1), 2629672. https://doi.org/10.1080/2331186X.2026.2629672

Kes, L., Wesel, F. V., Saunders-Smits, G. N., & Specht, M. M. (2025). *Defining Valid Criteria for Transversal Competency Assessment in Vocational Education and Training for Aircraft Maintenance using the Delphi Method*. In Review. https://doi.org/10.21203/rs.3.rs-6429387/v2

Krause, J., Portolese Dias, L., & Schedler, C. (2015). Competency-Based Education: A Framework for Measuring Quality Courses. *Online Journal of Distance Learning Administration*, *18*, 1–9.

Lerchenfeldt, S., Mi, M., & Eng, M. (2019). The utilization of peer feedback during collaborative learning in undergraduate medical education: A systematic review. *BMC Medical Education*, *19*(1), 321. https://doi.org/10.1186/s12909-019-1755-z

Lockyer, J., Carraccio, C., Chan, M.-K., Hart, D., Smee, S., Touchie, C., Holmboe, E. S., & Frank, J. R. (2017). Core principles of assessment in competency-based medical education. *Medical Teacher*, *39*(6), Article 6. https://doi.org/10.1080/0142159X.2017.1315082

McClelland, D. C. (1973). Testing for competence rather than for 'Intelligence'. *American Psychologist*, *28*(1), 1–14.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (Fourth edition). John Wiley & Sons.

Ndoye, A. (2017). Peer / Self Assessment and Student Learning. *International Journal of Teaching and Learning in Higher Education*, *29*(2), 255–269.

Norcini, J., Anderson, B., Bollela, V., Burch, V., Costa, M. J., Duvivier, R., Galbraith, R., Hays, R., Kent, A., Perrott, V., & Roberts, T. (2011). Criteria for good assessment:
Consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher*, *33*(3), 206–214. https://doi.org/10.3109/0142159X.2011.551559

Piggot-Irvine, E., Rowe, W., & Ferkins, L. (2015). Conceptualizing indicator domains for evaluating action research. *Educational Action Research*, *23*(4), 545–566.
https://doi.org/10.1080/09650792.2015.1042984

Prescott, L. E., Norcini, J. J., McKinlay, P., & Rennie, J. S. (2002). Facing the challenges of competency-based assessment of postgraduate dental training: Longitudinal Evaluation of Performance (LEP): Facing the challenges of a competency-based assessment system. *Medical Education*, *36*(1), 92–97. https://doi.org/10.1046/j.1365-2923.2002.01099.x

Resta, P., & Lee, H. (2010). Peer and Self-Assessment. In R. Ubell (Ed.), *Virtual Teamwork* (1st edn, pp. 45–64). Wiley. https://doi.org/10.1002/9780470615782.ch3

Schön, D. A. (2013). *The reflective practitioner: How professionals think in action*. Ashgate.

Schuwirth, L. W. T., & Van der Vleuten, C. P. M. (2011). Programmatic assessment: From assessment of learning to assessment for learning. *Medical Teacher*, *33*(6), 478–485. https://doi.org/10.3109/0142159X.2011.565828

Seifert, T., & Feliks, O. (2019). Online self-assessment and peer-assessment as a tool to enhance student-teachers' assessment skills. *Assessment & Evaluation in Higher Education*, *44*(2), 169–185. https://doi.org/10.1080/02602938.2018.1487023

Sluijsmans, D. M. A., Straetmans, G. J. J. M., & Van Merriënboer, J. J. G. (2008). Integrating authentic assessment with competence-based learning in vocational education: The Protocol Portfolio Scoring. *Journal of Vocational Education & Training*, *60*(2), 159–172. https://doi.org/10.1080/13636820802042438

Somekh, B. (1995). The Contribution of Action Research to Development in Social Endeavours: A position paper on action research methodology. *British Educational Research Journal*, *21*(3), 339–355. https://doi.org/10.1080/0141192950210307
Spencer, L. M., & Spencer, S. M. (1993). *Competence at work: Models for superior performance*. Wiley.

Strom, P. S., & Strom, R. D. (2011). Teamwork skills assessment for cooperative learning. *Educational Research and Evaluation*, *17*(4), 233–251. https://doi.org/10.1080/13803611.2011.620345

To, J., & Panadero, E. (2019). Peer assessment effects on the self-assessment process of first-year undergraduates. *Assessment & Evaluation in Higher Education*, *44*(6), 920–932. https://doi.org/10.1080/02602938.2018.1548559

Topping, K. J. (2017). Peer Assessment: Learning by Judging and Discussing the Work of Other Learners. *Interdisciplinary Education and Psychology*, *1*(1). https://doi.org/10.31532/InterdiscipEducPsychol.1.1.007

Tornwall, J. (2018). Peer assessment practices in nurse education: An integrative review. *Nurse Education Today*, *71*, 266–275. https://doi.org/10.1016/j.nedt.2018.09.017
Transport Canada. (2003). *HUMAN PERFORMANCE FACTORS FOR ELEMENTARY WORK AND SERVICING*.

Tremblay, M. C., Hevner, A. R., & Berndt, D. J. (2010). Focus Groups for Artifact Refinement and Evaluation in Design Research. *Communications of the Association for Information Systems*, *26*. https://doi.org/10.17705/1CAIS.02627

Van Den Eertwegh, V., Van Dalen, J., Van Dulmen, S., Van der Vleuten, C. P. M., & Scherpbier, A. (2014). Residents' perceived barriers to communication skills learning: Comparing two medical working contexts in postgraduate training. *Patient Education and Counseling*, *95*(1), 91–97. https://doi.org/10.1016/j.pec.2014.01.002

Van Der Vleuten, C., Lindemann, I., & Lisa Schmidt. (2018). Programmatic assessment: The process, rationale and evidence for modern evaluation approaches in medical education. *Medical Journal of Australia*, *209*(9), 386–388. https://doi.org/10.5694/mja17.00926

Van der Vleuten, C. P. M. (1996). The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education*, *1*, 41–67.

Van der Vleuten, C. P. M., Van Den Eertwegh, V., & Giroldi, E. (2019). Assessment of communication skills. *Patient Education and Counseling*, *102*(11), 2110–2113. https://doi.org/10.1016/j.pec.2019.07.007

Van Helden, G., Van Der Werf, V., Saunders-Smits, G. N., & Specht, M. M. (2023). The Use of Digital Peer Assessment in Higher Education—An Umbrella Review of Literature. *IEEE Access*, *11*, 22948–22960. https://doi.org/10.1109/ACCESS.2023.3252914

Villardón-Gallego, L., Flores-Moncada, L., Yáñez-Marquina, L., & García-Montero, R. (2020). Best Practices in the Development of Transversal Competences among Youths in Vulnerable Situations. *Education Sciences*, *10*(9), Article 9. https://doi.org/10.3390/educsci10090230

Winch, C., & Addis, M. (2022). Autonomy and expertise in the English workplace. *Journal of Vocational Education & Training*, *74*(1), Article 1. https://doi.org/10.1080/13636820.2020.1869808

Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Zylawski, C., & Ma, M. (2023). CBTA as a Solution to the MTX Talent Crisis. *Aviation Voices: Commercial Aviation Safety & Training*. https://aviationvoices.com/maintainer-competency-based-training-and-assessment-cbta-learning-from-air-france-industries/