

Challenges and Barriers in Vocational Learning Assessment: A Case Study of Automotive Engineering Expertise Program in Vocational High Schools

Suhartanta^{1,*}, Nirmala Adhi Yoga Pambayun¹, Fajrin Sidiq Muzaffarul Zaman¹, & Yusri bin Kamin²

¹Department of Automotive Engineering Education, Universitas Negeri Yogyakarta, Indonesia

²Department of Advanced Technical and Vocational Education and Training, Universiti Teknologi Malaysia, Malaysia

*Corresponding email: suhartanta@uny.ac.id

Received: 28 April 2026

Accepted: 30 June 2026

Published: 08 July 2026

Abstract: This study aims to analyze the challenges and barriers encountered in vocational learning assessment and to propose a model to strengthen assessment practices based on empirical findings. A qualitative multiple-case study was conducted in one public and one private vocational high school in the Special Region of Yogyakarta, Indonesia, from March to October 2025. Six productive-subject teachers with at least 8 years of teaching experience were purposively selected. Data were collected through in-depth semi-structured interviews and document analysis of lesson plans, teaching modules, assessment instruments, and learning outcome records. We combined document analysis with in-depth interviews (triangulation) to maximize data validity. Thematic analysis was carried out with ATLAS.ti 22 through data familiarization, initial coding, theme generation, theme review, theme definition, and reporting. Trustworthiness was strengthened through source triangulation, member checking, peer debriefing, and cross-case comparison. The study identified multidimensional challenges in the cognitive, psychomotor, and affective domains. Cognitive assessment was constrained by low student literacy engagement, the dominance of multiple-choice tests, limited use of diagnostic assessment, and weak follow-up of item analysis. Psychomotor assessment was constrained by limited facilities, limited teaching time, equipment wear, and uneven opportunities for observing individual performance. These barriers differed between the two schools due to variations in management flexibility, team teaching, procurement, and industry involvement. Affective assessment was constrained by large class sizes, limited observation time, teacher reliance on informal impressions, and the absence of explicit attitude assessment rubrics. The findings indicate the need for a context-sensitive model that strengthens meaningful assessment, continuous diagnostic-formative-summative cycles, resource collaboration, and teacher capacity. The proposed model remains hypothetical and requires further validation and testing of its effectiveness.

Keywords: vocational learning assessment, cognitive, psychomotor, affective, case study.

Article's DOI: <https://doi.org/10.23960/jpp.v16i3.pp1371-1394>

■ INTRODUCTION

Vocational High Schools (VHS) play a strategic role in preparing graduates for the workforce, especially in technical fields that require occupational competence. In the context of automotive engineering education, graduates are expected not only to understand theoretical concepts but also to perform practical work

procedures, apply safety standards, solve technical problems, and demonstrate professional attitudes in workshop activities. Therefore, vocational learning is closely related to the integration of cognitive, psychomotor, and affective competencies. This characteristic assesses vocational high schools differently from assessment in general academic education

because teachers must measure students' knowledge, practical skills, and work-related attitudes in an integrated manner. Previous studies have emphasized that vocational education should connect school-based learning with workplace standards and employability needs (Baartman et al., 2007; Gulikers et al., 2004; Suhartanta et al., 2023). In Indonesia, vocational high school graduates are also positioned to fill operational technical work at level II of the Indonesian National Qualifications Framework. Consequently, the quality of vocational education depends not only on curriculum implementation and teaching strategies but also on the extent to which assessment practices can validly represent students' readiness to perform work-related competencies.

Vocational high schools currently face several challenges related to the quality and relevance of graduate competencies to workplace demands. The Indonesian government has made various efforts to improve vocational education through curriculum renewal, industry collaboration, and strengthening of practical learning. However, improvement efforts often focus more on curriculum content and learning activities, while assessment practices have not received the same level of systematic attention. In this study, the statement that assessment has not been properly addressed does not mean that teachers do not conduct assessments.

Teachers already conduct written tests, practical tests, assignments, final semester assessments, and attitude observations. The main issue lies in the quality, continuity, and utilization of assessment results. Cognitive assessment is often dominated by written tests and is not always followed by diagnostic interpretation for improving instruction. Psychomotor assessment is constrained by limited equipment and time, and by the difficulty of observing each student's performance individually. Affective assessment is often based on informal teacher impressions because structured attitude assessment

instruments are not yet consistently used. These conditions are important to examine because assessment should serve not only as a grading mechanism but also as a source of feedback to improve learning, strengthen competence, and support students' readiness for the workplace (Black & Wiliam, 1998; Carless, 2007; Wiliam, 2011).

On the other hand, Ordoñez-Avila et al. (2023), Tosuncuoglu (2018), and Thanabalan et al. (2015) outline several objectives of assessment, including diagnosing student strengths and weaknesses, monitoring student progress toward achieving goals, assigning grades, determining instructional effectiveness, providing feedback, preparing students for high-stakes exams, and motivating students. Based on this statement, teachers must be able to define the objectives of the assessments to be implemented so that the assessment process aligns with those objectives and the results answer whether the objectives have been achieved. Assessment is used not only to determine student learning outcomes but also as part of the learning process (Mardapi, 2017). A good assessment will positively impact students' learning. In other words, students will be more motivated and have accurate reflection material to improve their abilities or competencies.

Assessment in vocational schools can be divided into three domains: cognitive, psychomotor, and affective (Abdurrahman & Mahmudah, 2023). Cognitive assessments in vocational schools often struggle to develop questions that align with industry needs. Many cognitive assessments remain theory-based, failing to connect to real-world situations. The lack of Higher Order Thinking Skills (HOTS)-based questions also hinders the measurement of students' analytical abilities (Lin et al., 2024; Pratama et al., 2022). Artificial intelligence-based adaptive testing technology is being developed to address this limitation (Lin et al., 2024). Psychomotor assessments in Automotive

Engineering require hands-on practice, which is often hampered by limited facilities and tools (Setiawan & Islami, 2020). Furthermore, skills assessment standards often vary across schools, hindering the integration of competency standards (Mustapha et al., 2020). Assessment of affective aspects such as discipline, responsibility, and work ethic is often overlooked or conducted subjectively by teachers (Wagner et al., 2024). Furthermore, the measurement of soft skills, a key factor in VHS graduates' success in the workforce, remains suboptimal.

Assessments used in learning can be categorized into two types: formative and summative (Mardapi, 2017). Formative assessment is the process of collecting evidence of student learning, providing feedback to students, and adjusting instructional and learning strategies to improve achievement (Chappuis & Stiggins, 2002). Based on this opinion, formative assessment is not merely an assessment of students' learning achievement but also serves to improve learning so that it better aligns with students' needs and the achievement of predetermined learning objectives. This view is consistent with Wiliam (2011), who emphasizes that assessment information should be used to support subsequent instructional decisions. Regarding the timing of assessment, Mardapi (2008) states that formative assessment is an integral part of the learning process. Therefore, this assessment is conducted concurrently with the ongoing learning process. Summative assessment is conducted when all material is deemed complete to determine grades based on students' learning outcomes (Ayub et al., 2020). In other words, teachers use summative assessment to determine students' latest developmental milestones in their mastery of competency standards.

However, issues related to assessment in vocational education, particularly in VHS, have not been thoroughly mapped and identified, including the impacts experienced by students

from teacher-administered assessments. Rezkilaturahmi et al. (2024); Rezkilaturahmi et al. (2025); Nisa & Retnawati (2018) explain that the purpose of assessment is not only to determine student learning achievement but also to provide feedback and motivate students to improve. Based on research by Pambayun et al. (2023), not all aspects of learning are assessed, indicating the continued weakness of assessment in VHS, particularly in automotive engineering programs. Besides the implementation of assessments on essential aspects of the education and learning process in vocational schools, there are still many aspects that have not been explored through research, such as the quality and completeness of assessment instruments used by teachers, assessment methodology, accuracy of assessment results, utilization of assessment results to improve the learning process, and feedback for students (Firmania et al., 2018). Therefore, research related to the problems of vocational learning assessment in VHS is necessary. The results of this research can serve as a basis for determining actions or innovations needed to strengthen vocational learning assessment in VHS.

Thus, this study aimed to determine the challenges and obstacles faced by teachers in implementing vocational learning assessments in the cognitive, psychomotor, and affective domains and to investigate a relevant and contextual vocational learning assessment strengthening model that can be formulated based on empirical findings of formative and summative assessment challenges in Automotive Engineering of VHS. Therefore, the research questions can be formulated as follows:

1. What are the problems of vocational learning assessment in vocational high schools in the cognitive, psychomotor, and affective domains?
2. What are the problems of formative and summative assessment in vocational learning in vocational high schools?

3. What is the hypothetical model for strengthening vocational education assessment based on the research findings?

■ METHOD

Participants

The participants of this study were six teachers from the automotive engineering expertise program. They came from two vocational schools in the Special Region of Yogyakarta, one public VHS and one private VHS. Both schools had A accreditation, adequate educational resources, a solid track record of student achievements, and good industry partnerships within the region. For these reasons, the findings from the public VHS and one private VHS in this study might still serve as a reference for best practices, or at least provide a picture of the kinds of challenges commonly faced by vocational schools. However, they are not intended to be broadly generalized.

The inclusion criteria were productive subject teachers in the Automotive Engineering Expertise Program who were actively teaching

and had at least eight years of experience in vocational education. These criteria were used to ensure that participants had sufficient knowledge of cognitive, psychomotor, affective, formative, and summative assessment practices in workshop-based learning.

At the public school, the participants included one teacher, who also served as head of the automotive engineering department, and two senior teachers. At the private school, the participants consisted of one teacher who also served as the vice principal for curriculum, one department head, and one senior teacher. Their teaching experience and school positions enabled them to provide rich information about assessment practices, school policies, facilities, and industry involvement in the Automotive Engineering Expertise Program. A demographic breakdown of the participants is presented in Table 1.

Research Design and Procedure

This study employs a multiple-case study design (Yin, 2018) with a holistic approach. Case

Table 1. Demographic of participants

No.	VHS	Number of Participants	Participants Code	Teaching Experience	Teacher's Field of Expertise
1	Public VHS	3	Teacher A	8.5 years	Chassis & Powertrain
			Teacher B	15 years	Engine
			Teacher C	17 years	Electricity
2	Privat VHS	3	Teacher D	10 years	Engine
			Teacher E	12 years	Electricity
			Teacher F	14 years	Chassis & Powertrain

studies are considered appropriate because they allow for an in-depth examination of complex assessment practices in the context of vocational learning in VHS, where assessment is closely linked to teaching, facilities, and industry demands.

Furthermore, this research was conducted in both public and private vocational high schools that offer Automotive Engineering Expertise Programs in the Special Region of Yogyakarta,

Indonesia. This multiple-case design allows for cross-case analysis to identify these two types of institutions, which were purposefully selected to capture variations in assessment implementation related to differences in governance, resource availability, management flexibility, and industry collaboration.

This study employed a multiple-case study design to explore in depth the challenges and barriers of vocational learning assessment in the

real-world context of two vocational high schools (one public and one private) with the automotive engineering expertise program. The case study allows the researcher to understand the phenomenon as a bounded system (Creswell, 2007).

The study was carried out from March to October 2025. The research procedure consisted of four stages: (1) Preparation stage, semi-structured interview guidelines were developed based on assessment and vocational education literature, and then validated by experts; (2) Data collection stage, in-depth interviews were conducted with the six participants, along with document analysis (lesson plans, assessment instruments, and learning outcome records) and supporting observations; (3) Data analysis stage, thematic analysis was performed using ATLAS.ti 22, followed by a cross-case analysis; (4)

Reporting stage, the findings were compiled and integrated with the literature.

Instruments

Data were collected through in-depth, semi-structured interviews and document analysis, enabling triangulation and greater trustworthiness. Therefore, in-depth interviews and semi-structured interview guidelines were developed based on assessment theory and vocational education literature. The document analysis sheet was developed by the researchers to review the lesson plans and the assessment instruments. The validity of the instrument was examined through expert judgment from two lecturers specializing in vocational education evaluation and one senior teacher in automotive engineering. Interview questions focus on the table mentioned.

Table 2. Interview focus area

No.	Focus Area	Description of Interview Focus
1	Assessment challenges	Teachers' experiences and difficulties in assessing cognitive, psychomotor, and affective domains in vocational learning.
2	Formative and summative assessment practices	Implementation strategies, frequency, and constraints of formative and summative assessments, including diagnostic and remedial practices.
3	Assessment instruments and digital platforms	Types of assessment instruments used (tests, rubrics, and observation sheets) and the utilization of digital platforms such as CBT or school-based assessment systems.
4	Use of assessment results	How assessment results are analyzed, interpreted, and used for learning improvement, feedback, and instructional decision-making.
5	Industry relevance in assessment	Teachers' perceptions of industry alignment, including industry partners' involvement and the relevance of assessment standards to workplace demands.

Data Analysis

Thematic analysis was conducted using ATLAS.ti 22 software and followed the stages proposed by Braun & Clarke (2006): (1) data familiarization, through repeated reading of interview transcripts and documents; (2) initial coding, identifying meaningful units related to assessment challenges and practices; (3) theme generation, grouping codes into preliminary themes; (4) theme review, examining the

coherence and relevance of themes across cases; (5) theme definition and naming, refining themes to clearly represent the core findings; (6) report writing, integrating themes with empirical evidence and relevant literature. Thus, cross-case analysis was conducted to identify similarities and differences between public and private vocational high schools.

To ensure the trustworthiness of the findings, several strategies were applied.

Triangulation was done first by comparing data from different sources, interviews, document analysis, and field notes, and also across participant groups (teachers from public and private schools). Member checking came next; preliminary findings were shared with all participants to see whether the interpretations matched their actual experiences. Their feedback was then worked into the final refinement of themes. Peer debriefing was also conducted through regular discussions with two colleagues who were not involved in data collection but had expertise in vocational education research. Taken together, these steps helped keep the thematic analysis credible, dependable, and confirmable.

■ **RESULT AND DISCUSSION**

Learning Assessment's Problems In The Cognitive, Psychomotor, And Affective Domains

Vocational learning in vocational high schools encompasses both theory and practice, with the latter being a dominant component. Therefore, assessment should ideally be conducted comprehensively across all competency domains: cognitive, psychomotor, and affective, each with its own characteristics and challenges. The findings from the six participants suggest that, in this study, teachers in both public and private schools faced relatively similar challenges in the cognitive domain. However, due to the limited number of participants, these results are not intended to be generalized. Rather, they provide preliminary insight into the assessment practices of the specific teachers who took part. One of the main challenges is students' low literacy and critical-thinking skills regarding theoretical material. Students tend to be more enthusiastic about practical learning than theoretical learning, resulting in cognitive test results often falling below the Minimum Completion Criteria (KKM), even though the material has been taught, and the

difficulty level of the questions has been adjusted to the basic to intermediate cognitive levels (C1–C3) as the participant mentioned that "...in learning, students often have difficulty understanding how components work and their functions, even though the material has been taught. After tests, many students still cannot recognize the components they should have learned. Based on our observations, student engagement in learning, especially in literacy, remains very low. Schools provide reading materials and learning resources, but most are not used by students. Student literacy tends to stall if not directly encouraged or guided by teachers. Therefore, we try to combine the use of books with digital devices such as cell phones, although cell phones can also be a barrier because they often disrupt focus during both theory and practice. Strengthening literacy is key because when students have strong literacy skills, their understanding of productive and practical learning improves. In practice, some students do look for tutorials on YouTube when they feel the teacher's explanation or textbooks are insufficient, but this has not yet become a systematic learning habit".

The quotation above indicates that literacy is not a separate academic issue in automotive vocational learning, but a prerequisite for competence in the workshop and workplace (Misbah et al., 2020). Automotive students are not only expected to operate tools or memorize component names; they must also be able to read service manuals, job sheets, maintenance schedules, wiring diagrams, diagnostic flowcharts, connector codes, torque specifications, and standard operating procedures. When technical literacy is weak, students may imitate the procedures demonstrated by the teacher but have difficulty explaining why a component fails, why a measurement is outside specification, or which troubleshooting step to carry out next. Consequently, low literacy becomes a barrier to cognitive assessment because questions about

component functions, system relationships, and troubleshooting sequences require students to connect written information with mechanical and electrical principles. It also affects psychomotor assessment because practical tasks usually begin with reading work instructions, identifying service specifications, and following safety procedures before touching the vehicle or component. This explains why the teachers identified literacy as a central problem: weak literacy reduces students' ability to move from knowing to interpreting to performing. Therefore, the next problem related to the dominance of written multiple-choice tests becomes more critical because such instruments may measure recognition of terms. However, they do not fully show whether students can interpret technical information as required in real automotive work. In contrast, authentic vocational assessment should represent real tasks, work contexts, outputs, and criteria (Gulikers et al., 2004).

Furthermore, teachers consider the theoretical material in the vocational high school curriculum complex and misaligned with the basic needs of the workforce at the entry-level technician level. Another problem concerns the limitations of cognitive assessment methods, which remain dominated by multiple-choice tests. This information from the teachers was also confirmed by the results of document analysis on the lesson plans and assessment instruments, which showed that most of the theoretical learning assessments used multiple-choice tests (80%). While this method facilitates the correction process, especially with the support of digital platforms such as G-School or Computer-Based Tests (CBT), these types of tests are considered incapable of optimally measuring students' analytical abilities. The use of essay tests or alternative forms of assessment is rarely implemented because they are considered to require more time for correction. Furthermore, diagnostic assessments, such as pre-tests or

formative assessments to map students' initial understanding, have not been conducted systematically and continuously. Teachers tend to rely on general observations and short quizzes, without in-depth analysis, as the basis for improving learning. At private VHS, this problem is exacerbated by the lack of routine analysis of question validity and reliability, raising concerns about the repeated use of the same questions year after year without ensuring item quality. On the other hand, the loss of the productive theory of the National Exam (UN) is also perceived to have impacted students' sense of urgency and seriousness in facing cognitive tests. Although public vocational high schools have been equipped with the G-School platform, which provides automatic item analysis, teachers have not optimally utilized these results to improve the quality of assessment instruments. The lack of systematic follow-up on cognitive assessment results likely stems from several interrelated causes. Heavy teaching loads and administrative responsibilities limit teachers' time (Black & Wiliam, 1998). Some teachers may also lack the competence to analyze assessment data and turn it into actionable teaching strategies (Heritage, 2010). In addition, schools often provide no clear policies or regulatory support to guide or require such follow-up (Fullan, 2015). Therefore, addressing this condition requires more than just calling for policies; it demands practical solutions such as reducing time burdens, offering data-literacy training, and building a supportive school culture that prioritizes assessment-driven improvement.

Meanwhile, regarding the psychomotor domain, several structural and technical challenges were identified. The primary issue lies in limited facilities and practical infrastructure. Practical tools and materials are expensive, in limited supply, and require a lengthy procurement process because they must be planned well in advance. One teacher stated, "...practical tools

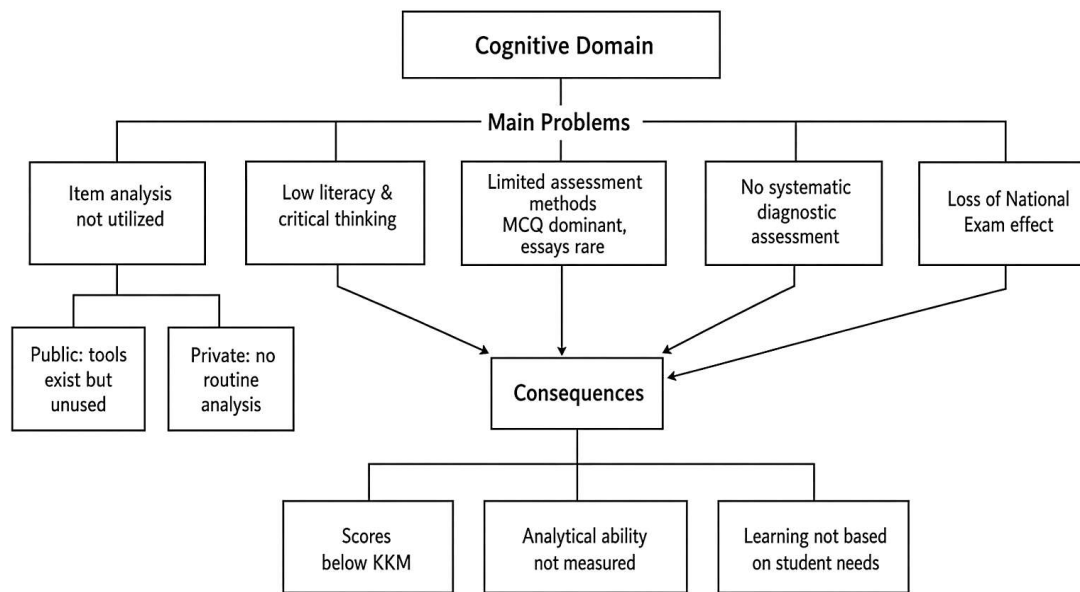


Figure 1. Thematic analysis results on assessment challenges in the cognitive domain

and materials are expensive and limited in quantity, and the procurement process takes a long time, so they cannot meet the needs of practical exams.” This results in many students having to share training equipment. In other words, a single piece of training equipment is used by multiple students. This accelerates equipment wear and tear and leads to inconsistent conditions during practical assessments. This situation was emphasized by teachers who noted that a single piece of equipment is often used by multiple students, resulting in varying equipment conditions during assessments. This inconsistency directly reduces the objectivity of psychomotor assessment results. Furthermore, time constraints pose significant challenges due to the large number of students and the diverse competencies to be assessed. Individual assessments are time-consuming, and public vocational school policies do not permit team teaching, requiring a single teacher to manage both teaching and assessment simultaneously. One respondent explained, “...with many students in one class and the absence of a team teaching policy, it is difficult to conduct objective individual practical assessments within the limited time”. This combination of limited facilities, time, and human resources

contributes to subjectivity in practical assessments.

In response to these limitations, teachers in public vocational high schools implemented an integrated practical assessment model. Specifically, several productive subjects are combined into a single, centrally managed practical assessment activity before the end-of-semester assessment, such as integrating competencies in engine adjustment and braking systems. One teacher explained that “...integrated practical assessment allows multiple competencies to be assessed simultaneously within a single practical theme, making the assessment process more efficient.” This model is considered more efficient in terms of time and resource utilization, increases student motivation, and better represents actual industrial work conditions. Furthermore, the integrated assessment model allows for greater teacher involvement in the assessment process, thereby increasing the objectivity and accuracy of psychomotor assessments.

The findings suggest that, within the two schools studied, psychomotor assessment challenges were reported as relatively more manageable by teachers at the private school, possibly due to greater management flexibility

observed there. That said, this should not be read as a rigid public-versus-private dichotomy. The interpretation here is therefore limited to the specific cases and should not be generalized into stereotypes about all public or all private vocational high schools.

This flexibility at the private school includes implementing team teaching, procuring facilities and equipment, and developing strong industry partnerships. One teacher stated, "...we have more flexibility in implementing team teaching and procuring equipment, which helps practical assessments run more effectively." A private vocational high school, one of the largest in the Special Region of Yogyakarta, has established strong partnerships with industry, particularly with Daihatsu. An industry-based curriculum is implemented not only in learning activities but also in assessments. As noted by a respondent, "...practical assessments in industrial classes follow Daihatsu's curriculum standards, and industry personnel are directly involved as examiners", ensuring the relevance and validity of psychomotor assessment to workplace needs.

Despite this, private vocational high schools still face several challenges in implementing psychomotor assessments, particularly related to student discipline during practical sessions. Student absences necessitate teachers providing supplementary practical assessments, which require time, effort, and access to additional equipment and materials. One teacher explained, "...when students are absent during practical sessions, we have to conduct supplementary assessments, which requires additional time and equipment preparation." Furthermore, the high frequency of equipment use among many

students accelerates wear and tear. One respondent stated that "...vehicle components are not designed for repeated disassembly and reassembly, so frequent use leads to faster functional deterioration." However, these challenges remain manageable thanks to flexible procurement policies and adequate funding support from foundations government assistance.

The strengths of private vocational high schools in supporting optimal psychomotor assessment are reflected in their strategic equipment management, structured block systems, and effective implementation of team teaching. Responsive procurement policies enable prompt repair or replacement of damaged equipment, while using machines with longer lifespans helps maintain the condition of newer equipment. However, this should not be read as a rigid dichotomy between public and private schools. The block system allows students to focus intensively on workshop activities over several days, with team teaching by multiple teachers. Strong industry support further enhances the quality of assessment through curriculum alignment and direct industry involvement in competency certification exams. These findings highlight that strong collaboration between VHS and industry is crucial to ensuring the relevance, objectivity, and quality of psychomotor assessments.

To clarify the cross-case evidence, Table 3 summarizes how industry partnership practices differed between public and private vocational high schools in assessment planning, external assessment, equipment procurement, team teaching, and implications for psychomotor assessment quality.

Table 3. Comparison of industry partnership practices in public and private VHS

No	Aspect	Public VHS	Private VHS
1	Industry involvement in assessment planning	Industry input existed but was not systematically integrated into daiy assessment instruments. Planning was mostly teacher-led and based on school curriculum documents.	Industry involvement was more explicit through industrial class standards, especially the partnership with Daihatsu.

2	Role as external assessor	External assessors were limited to selected competency activities and were not consistently involved in routine practical assessment.	Industry personnel were directly involved as external examiners in practical assessment and competency certification.
3	Alignment with workplace standards	Assessment standards were adapted to available facilities and learning time, so workplace simulation was not always fully represented.	Assessment standards followed industrial class procedures, making workplace relevance stronger.
4	Flexibility of equipment procurement	Procurement required longer administrative procedures and advance planning, which affected the facility's readiness for practical assessment.	Procurement and maintenance were more flexible due to foundation support, management flexibility, and industry linkage.
5	Team teaching and assessment support	Team teaching was limited, so one teacher often handled instruction, supervision, safety control, and assessment simultaneously.	Team teaching was more flexible, allowing several teachers to support supervision and assessment of workshops.
6	Implication for assessment quality	Assessment faced stronger constraints in objectivity, time allocation, and facility readiness.	Assessment was more manageable because of industry standards, teacher collaboration, and equipment support strengthened the validity of practical assessment.

Thematic analysis results on assessment challenges in the psychomotor domain are depicted in the following figure:

Finally, the findings relate to the affective domain. Challenges are particularly evident related to limited in-depth observations of student

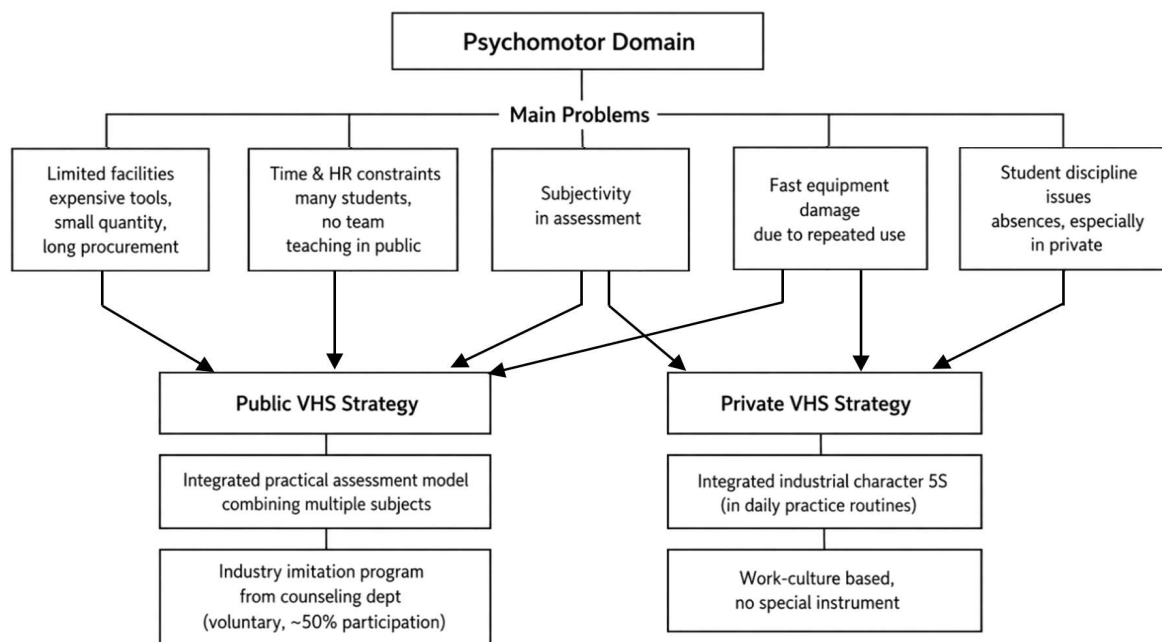


Figure 2. Thematic analysis results on assessment challenges in the psychomotor domain

attitudes. With relatively large class sizes and limited learning time, teachers struggle to comprehensively observe affective aspects such

as motivation, responsibility, and teamwork skills in each student. One teacher explained, "... the number of students in one class is quite large,

while learning time is limited, so it is difficult to observe the attitude of each student in depth, especially if their attitude is not very prominent ...”.

The teacher-student ratio provides a concrete explanation for why observing affective competence was difficult to conduct in depth. The teachers’ statement about large class size should not be understood solely as a general complaint about the number of students, but rather as a practical limitation on the collection of valid attitude evidence during workshop learning. Based on class information collected during the study, one teacher generally supervised 32 to 36 students per learning session. In the public vocational school, the ratio during regular workshop assessments was approximately 1:36 when team teaching was unavailable. In contrast, at the private vocational school, the ratio ranged from approximately 1:32 to 1:34, depending on the learning block and teacher availability. With a 90-120 minute learning session, this ratio leaves only a very short observation window for each student, especially when the teacher must also explain tasks, supervise safety, check tools, and assess work results. As a result, subtle affective behaviors such as consistency, initiative, cooperation, responsibility, and sustained work discipline are harder to record than visible incidents such as absence, lateness, unsafe behavior, or disruption. This ratio also explains why teachers tend to remember students with very positive or very negative behavior more easily than students with moderate attitudes. Thus, the quantitative ratio data strengthens the following analysis of observational bias and the possibility of the halo effect in affective assessment. To address this problem, structured observation rubrics and peer assessment are important because they make evidence collection more transparent and help students assess responsibility, cooperation, and contribution during practical work (Zhang & Hwang, 2023).

The tendency to capture only extreme positive or negative behaviors while missing moderate ones can be explained by several factors. In large, crowded classes, teachers’ attention is naturally drawn to highly visible incidents, major disruptions, or outstanding acts because those demand immediate responses (Airasian, 2001). Meanwhile, moderate level attitudes rarely stand out. Cognitive biases, particularly the *halo effect* (Thorndike, 1920), further skew judgments: a teacher’s overall impression of a student (based on academic or skill performance) often colors the assessment of that student’s attitude, compressing moderate variations. Additionally, affective assessment is inherently difficult because attitudes are internal and inconsistently displayed; without structured observation guides, teachers rely on memorable extremes rather than representative samples (McMillan, 2018). Thus, the finding is not simply that teachers assessed poorly, but that classroom conditions and natural cognitive shortcuts systematically favor extreme observations.

This condition was further emphasized by another teacher who noted that “... which are usually only seen by very active or problematic students, while attitudes such as responsibility and cooperation cannot always be observed comprehensively”. Furthermore, affective scores were not explicitly displayed on student report cards, limiting their impact on students’ awareness of the importance of attitude and character in the learning process. As one respondent stated, “... attitude scores are not specifically written on report cards and are used more as internal teacher notes.”

To address the aforementioned limitations, public vocational high schools have implemented several initiatives to strengthen and assess the affective domain. One such effort involves using peer assessment in group work, in which students evaluate their peers’ contributions, diligence, and responsibility. According to one teacher, “... we

tried peer assessment to see students' responsibilities and contributions during group work." However, this peer assessment remains relatively general, relying primarily on a rating system without detailed indicators or rubrics, which can increase subjectivity, as acknowledged by a teacher who stated that "... the assessment is still simple and does not use a detailed rubric." In addition, the school implemented a targeted affective-strengthening program through an "industry imitation" initiative coordinated by the Department of Guidance and Counseling. This program aims to develop work readiness, professional attitudes, communication skills, and preparation for industrial psychology tests. However, its effectiveness is limited due to its voluntary nature, with student participation reaching only about half of the total student population.

Meanwhile, findings from a private vocational high school indicate that affective domain assessment has not been implemented using a stand-alone instrument but has been integrated into daily practical learning activities and teacher observations. One teacher explained that "... we usually assess attitudes based on students' daily practices, such as discipline and responsibility. There is no specific instrument for the affective aspect." This situation is exacerbated by persistent basic discipline issues, such as inconsistent student attendance, including during practical sessions and exams. As one respondent noted, "... there are still students who are frequently absent, even during practical sessions and exams, making it difficult to assess their attitudes on an ongoing basis. This inconsistency makes it difficult for teachers to maintain continuity of learning and to obtain a comprehensive and ongoing picture of students' affective development."

In an effort to strengthen the affective domain, the private vocational high school implemented an integrated industrial character

development program across all learning activities. The school emphasized industrial cultural values, such as the application of the 5S/5R principles, work discipline, responsibility, and commitment to duty, as part of the daily practical learning routine. A teacher highlighted that "we always apply industrial cultural values such as 5S, work discipline, and responsibility in our practical activities because that is what is needed in the world of work". This approach not only strengthened students' character but also provided a contextual basis for assessing attitudes aligned with workplace demands, demonstrating the potential of integrating industrial cultural values into the sustainable development and assessment of the affective domain.

In this context, the 5S strategy clarifies how the private vocational school attempted to translate industrial character into observable daily behavior. 5S refers to Seiri, Seiton, Seiso, Seiketsu, and Shitsuke. Seiri means sorting and separating necessary tools, components, and materials from unnecessary ones. Seiton means arranging tools and components in the correct place so that they are easy to identify, use, and return. Seiso means cleaning the workplace, tools, and components after practice. Seiketsu means standardizing work procedures, tool placement, cleanliness routines, and safety habits. Shitsuke means maintaining discipline so that these habits are carried out consistently without repeated reminders from the teacher. In daily workshop assessment, these principles can be observed through concrete indicators, such as whether students prepare only the required tools before practice, return wrenches and special service tools to their proper locations, clean oil or dust after disassembly, label components correctly, dispose of used materials properly, use personal protective equipment, and maintain workshop order after completing the task. This makes 5S more than a slogan of industrial culture; it becomes a practical bridge between affective assessment

and psychomotor performance, as discipline, responsibility, safety awareness, and care for tools can be observed while students complete technical work. In line with this, 5S can support workplace organization, standardization, and sustainable work habits when implemented as a daily culture rather than a temporary program

(Randhawa & Ahuja, 2017). Therefore, the 5S-based routine provides a more contextual basis for assessing affective competence in automotive workshops. However, it still needs to be supported by explicit rubrics to make the evidence collected by teachers more consistent, transparent, and accountable.

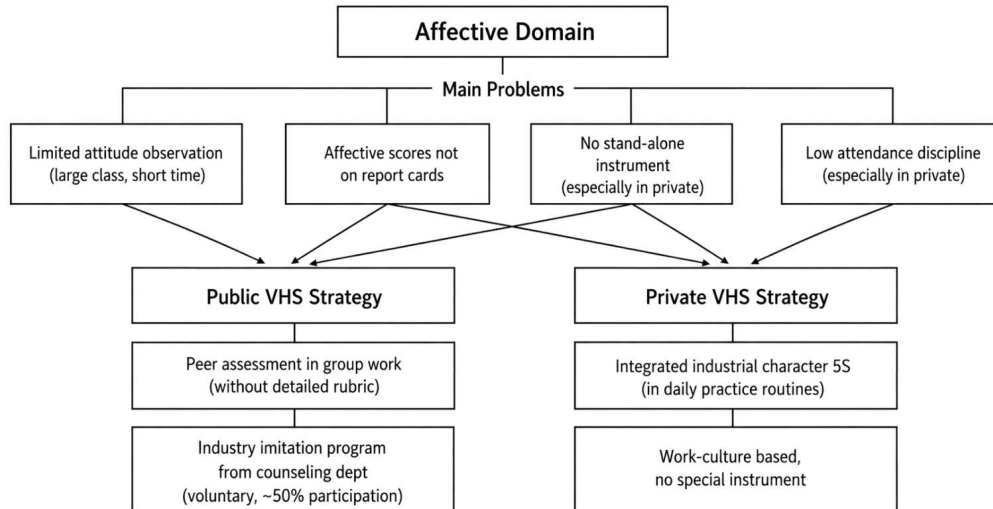


Figure 3. Thematic analysis results on assessment challenges in the affective domain.

Based on the research findings, the public and private vocational high schools encountered relatively similar obstacles in formative and summative assessment, as described below.

The findings indicate that formative assessment in vocational schools has not been implemented consistently and optimally. Formative assessment should function as a tool to monitor student learning progress and improve the learning process. However, its implementation is often hampered by time constraints. Teachers are required to complete dense learning materials, so formative assessment is frequently sidelined and not implemented systematically. Various school agendas outside regular classes, such as commemorations of major holidays, the Pancasila Student Profile Strengthening Project (P5), and other incidental programs, also disrupt the planned formative assessment schedule. As a result, formative assessment has not been used optimally

as assessment for learning and tends to be sporadic and unsustainable.

Problems in summative assessment primarily relate to the high administrative burden teachers must bear. Teachers feel that the numerous administrative demands of assessment are disproportionate to their professional judgment, even though they have interacted with students and come to understand their learning characteristics over time. Remedial programs for students who have not yet met the Minimum Mastery Criteria (KKM), both in theory and practice, remain limited. Remedial activities generally consist of retake exams without sufficient additional learning because of time constraints and busy learning schedules. Student absence exacerbates this situation, making it difficult for teachers to provide effective and sustainable remedial services. Although analysis of the validity and reliability of theory test

instruments has been facilitated through digital platforms such as G-School, item analysis has not been conducted routinely and comprehensively each semester. Teachers tend to reuse questions from previous years without

adequate quality evaluation, so the validity and reliability of summative assessment instruments are not fully guaranteed. A summary of the assessment practices in VHS identified in this research is presented in the following table.

Table 4. Summary of assessment findings in public and private VHS

No	Assessment Aspect	Finding in Public VHS	Finding in Public VHS
1	Cognitive assessment	Assessment was dominated by written and multiple-choice tests. G-school was available, but item analysis was not yet optimally used to improve learning and instruments.	Assessment was also dominated by multiple-choice tests; item validity and reliability were not routinely analyzed, and similar questions were often reused.
2	Technical literacy	Students had difficulty understanding component functions, system relationships, job sheets, service procedures, and troubleshooting tasks	Similar literacy challenges appeared when students connected theory with workshop procedures; online tutorials were used, but not systematically
3	Psychomotor assessment	Practical assessment was constrained by limited tools, workshop time, equipment wear, and the absence of flexible team teaching.	Practical assessment was more manageable due to block scheduling, flexible team teaching, procurement support, and stronger industry partnerships, although absences and equipment wear remained challenges.
4	Affective assessment	One teacher supervised approximately 36 students during a workshop session, so attitude assessment often relied on general impressions, peer assessments, and internal notes.	Affective assessment was integrated into daily workshop routines through discipline, responsibility, 5S/5R, and industrial character practices, with a ratio of approximately 1:32 to 1:34.
5	Formative assessment	Formative assessment was inconsistent due to dense material, school agendas, and limited time. Feedback and follow-up were not fully documented.	Similar constraints existed. Teachers used short quizzes and informal observation, but formative assessment had not become a continuous cycle.
6	Summative assessment	Summative assessment was influenced by administrative demands and final grading. Remedial activities were commonly limited to retake tests.	Summative assessment was also affected by grading pressure, but industrial classes had a stronger potential to connect assessment with competency standards and external assessment.
7	Assessment documents	Lesson plans, question sheets, and assessment instruments existed, but documentation of feedback, remedial learning, item analysis follow-up, and affective rubrics was limited.	Documents existed with stronger support from industry-based standards in selected classes, but affective instruments and feedback records still needed improvement.
8	Industry involvement	Industry involvement existed but was not systematically integrated into daily assessment planning and routine practical assessment.	Industry involvement was more explicit through Daihatsu-supported industrial classes, workplace-oriented standards, and external examiners.

9	Overall condition	Assessment was constrained by procurement bureaucracy, limited team teaching, large student groups, limited follow-up of digital item analysis, and difficulties in observing affective competence.	Assessment was supported by managerial flexibility, team teaching, block scheduling, industrial culture, and stronger industry linkage, although discipline and equipment wear remained challenges.
---	-------------------	---	---

Table 4 shows that the assessment problems in the two VHS were not completely different in substance, but differed in the institutional capacity to respond to them. Both schools faced low student literacy, dominance of multiple-choice tests, limited diagnostic and formative assessment, psychomotor constraints, and difficulties in assessing affective competence. However, the private VHS had greater flexibility in team teaching, equipment procurement, block scheduling, and industry-based assessment support. In contrast, the public VHS had digital assessment support and integrated practical assessment initiatives, but still faced stronger administrative and managerial constraints. These findings indicate that assessment strengthening in vocational education should be designed as a contextual model that combines teacher capacity, assessment instruments, facility management, school policy, and industry collaboration. Model for strengthening vocational education assessment

The model for strengthening vocational education assessment in this study was formulated based on an analysis of real-world problems in vocational high schools across the cognitive, psychomotor, and affective domains, as well as on the implementation of formative and summative assessments. In the cognitive domain, key issues include low student interest in reading and literacy, theoretical material perceived as uninteresting and overly complex, suboptimal diagnostic formative assessments, and the unassured validity and reliability of theory test instruments. Furthermore, the elimination of the national productive theory exam has reduced students' sense of urgency in facing academic assessments.

In the psychomotor domain, prominent issues include limited practical facilities and

infrastructure, particularly in public vocational high schools; limited assessment time to assess each student's skills individually; high levels of wear and tear on practical equipment; and a gap between the condition of school equipment and industry standards. Meanwhile, in the affective domain, teachers face difficulties with in-depth observation and assessment of student attitudes due to the large number of students, the lack of structured, explicit attitude assessment instruments in learning outcome reports, and the still-weak culture of discipline and responsibility among some students. Regarding formative and summative assessment, research findings indicate that formative assessment is often sacrificed due to the demands of completing material and school agendas, resulting in underutilization for improving learning. Summative assessment, on the other hand, is burdened by excessive administration and pressure to produce final grades, which in practice often neglects the process of validly and comprehensively demonstrating student competency. Thus, problem-solving strategies and efforts to strengthen vocational assessments can encompass three aspects: enhancing the meaningfulness of assessment, strengthening the assessment process, and strengthening the capacity of resources, teachers, and instruments.

Strengthening the meaningfulness of assessments

Strengthening the perceived value of assessments is achieved by rebranding them as "Competency Tests" or "Mini Certifications," which mirror industry selection and certification processes. Assessments are no longer seen merely as a means of assigning grades, but rather as a means of demonstrating competencies

relevant to job requirements. Concrete efforts undertaken include involving industry practitioners as guest examiners in the main summative assessment, designing assessments with scenarios and standards that resemble company entrance tests or Professional Certification Institutions (LSP), and conducting extensive outreach to students, explaining that each assessment is part of building a competency portfolio that will be assessed by industry.

Strengthening the assessment process

Strengthening the assessment process is achieved by implementing an integrated, continuous assessment cycle comprising diagnostic, formative, and summative assessments. Initial diagnostic assessments are conducted through theory pre-tests and observations of learning interests at the beginning of the material or learning block to map student needs. Continuous formative assessments are conducted through simple yet consistent techniques, such as exit tickets or short quizzes at the end of theoretical lessons, as well as practical observation checklists to assess work procedures, safety, and attitudes throughout the process. Digital platforms are utilized to support quick quizzes and simple item analysis. Furthermore, summative assessments are designed in an integrated, authentic manner, adopting a practical exam model that combines several productive subjects into a single project. Therefore, assessments focus not only on the final product but also on work processes, problem-solving, teamwork, and work attitudes based on industry culture.

Resource strengthening

Resource strengthening is carried out through collaborative strategies and optimization of facilities, time, and human resources. In facility management, schools are encouraged to create industrial workshop simulations with equipment

arranged by work area, implement a strict equipment loan and maintenance system, and leverage the flexibility of facility procurement in private vocational schools as best practices to advocate to the government for public vocational schools. In terms of time and human resource management, team-teaching policies are encouraged in practical learning to reduce the teacher-student ratio during assessments, and a dedicated “competency test” day is established in the academic calendar so that all resources can be focused on implementing summative assessments.

Strengthening teacher and instrument capacity

Strengthening teacher and instrument capacity is carried out through training and simplification strategies. Teachers need specific training in designing authentic assessments, developing attitude assessment instruments, and using simple item analysis techniques. Furthermore, developing a bank of analyzed theoretical questions and a bank of practical and attitude assessment instruments is a strategic step to ensure the consistency and quality of assessments. Attitude assessments are developed in a structured manner through rubrics with clear indicators, such as discipline and cooperation, and are integrated with peer assessments using the same criteria, thereby making affective assessments more objective, transparent, and accountable. A hypothetical model of strengthening vocational assessment in VHS is depicted in the following figure:

Field findings indicate low student interest and cognitive learning outcomes. This aligns with research by Ordoñez-Avila et al. (2023), which found that traditional teacher-centered pedagogical approaches and rote assessments are ineffective in vocational education contexts. Vocational high school students, who often have a kinesthetic learning orientation, require a meaningful learning context in which assessment

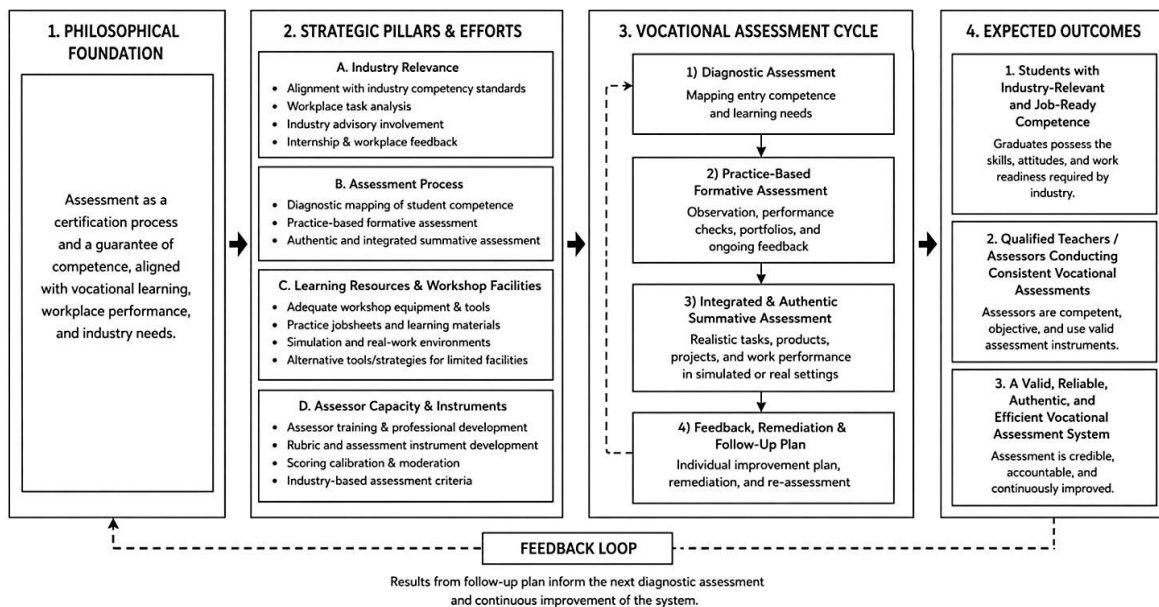


Figure 4. Hypothetical model for strengthening vocational assessment practices

is connected to concrete tasks, work processes, and practical performance. Therefore, assessment in vocational education should not be seen merely as a tool for measuring final achievement but as a learning mechanism that helps students identify competence gaps, reflect on their progress, and improve their performance. This finding is also consistent with Misbah et al. (2020), who showed that the implementation of competence-based vocational education in Indonesia still faces a gap between policy design and actual learning practices in schools.

The proposed hypothetical model addresses this issue by emphasizing meaningful assessment and ongoing formative assessment. This concept aligns with the principle of assessment for learning, in which assessment is designed to provide feedback and inform instructional decisions (Black & Wiliam, 1998; Wiliam, 2011). Techniques such as exit tickets, digital formative quizzes, pre-tests, and process observation checklists, as proposed in the model, not only monitor students' understanding but also train their metacognitive awareness. In this perspective, assessment information should be used to support subsequent instructional decisions rather than merely to produce scores.

However, the findings also reveal a critical contradiction between the ideal concept of assessment as learning and the actual conditions in vocational schools. In theory, assessment as learning requires continuous feedback, student reflection, authentic tasks, and active teacher facilitation. In practice, teachers often face limited instructional time, administrative workloads, limited workshop facilities, and different levels of student readiness. This contradiction indicates that the main problem is not simply the absence of formative assessment strategies, but the mismatch between the expected assessment practices and the institutional capacity available to support them. Therefore, strengthening vocational assessment should be understood as an assessment ecosystem that integrates teacher capacity, assessment instruments, workshop facilities, feedback mechanisms, and industry relevance.

For this reason, the proposed hypothetical model does not treat authentic assessment as a single final test. Instead, it positions vocational assessment as a continuous cycle that begins with diagnostic assessment, continues with practice-based formative assessment, and concludes with authentic summative assessment, feedback,

remediation, and reassessment. Theoretically, this model extends the concept of assessment as learning into the vocational education context. In vocational schools, assessment is not only a pedagogical process for improving learning but also a mechanism for ensuring competence, workplace readiness, and alignment with industry standards. Thus, vocational assessment needs to combine multiple methods, sources of evidence, and quality criteria to assess students' competence more comprehensively (Baartman et al., 2007).

Research by Abdurrahman & Mahmudah (2023) demonstrated that implementing assessment as learning significantly improved vocational high school students' engagement and conceptual understanding, as they directly identified gaps between their knowledge and expected standards. The infrastructure and time constraints expressed by teachers, particularly in public vocational high schools, are manifestations of classic challenges in performance assessment. According to Rezkilaturahmi (2024), authentic assessment requires tasks that replicate real-world challenges and standards. However, this often suffers from administrative complexity and resources.

Hypothetical models offer Integrated Practical Exams and problem-based projects (troubleshooting) as solutions. This approach is strongly supported by research by Gulikers et al., which states that authentic assessment in vocational education must have five dimensions: task, physical context, social context, outcome/form, and criteria (Pambayun et al., 2023). A motorcycle repair project that combines competencies from several subjects (as proposed in the model) meets all these dimensions: the task is authentic (finding and repairing damage), the context is in a workshop (physical), it is conducted in a team (social), the result is a functioning motorcycle (outcome/form), and the criteria can be adopted from industry standards. A study by Baartman (2008) on competence

assessment programs (CAPs) also showed that this type of integrated assessment has higher validity because it assesses competencies holistically, rather than piecemeal. Teachers' difficulties in objectively assessing attitudes across large student populations confirm Jiang et al. (2023) findings that affective assessment is often a "black box" due to its reliance on subjective impressions and the lack of structured instruments. The hypothetical model addresses this by proposing a structured Attitude and Peer Assessment Rubric. Furthermore, the integration of structured peer assessment, as proposed in the model, not only lightens teachers' burden but also fosters students' judgment and social responsibility, which are essential 21st-century competencies. A meta-analysis by Zhang & Hwang (2023) concluded that peer assessment has a significant positive effect on academic achievement and metacognitive skills.

The dominance of summative assessment and the weakness of formative assessment, as revealed in interviews, reflect a persistent culture of assessment of learning. Chappuis & Stiggins (2002), in their seminal work on formative assessment, assert that effective and timely feedback is at the heart of learning improvement. The hypothetical model places the Continuous Assessment Cycle at the core of the process. This cycle aligns with the concept of "closing the loop" in assessment, in which information from diagnostic and formative assessments is used to design learning interventions before summative assessment is administered (Pals et al., 2023). The use of technology (such as G-School) for rapid item analysis at Public Vocational High School can expedite the provision of this feedback. Research by Putra (2023) emphasizes that effective formative assessment must be instructionally tractable, meaning that teachers can immediately use the results to adjust their teaching. The proposed model explicitly encourages this through pretests process-observation checklists.

Findings on the effectiveness of industry partnerships at private vocational high schools and on bureaucratic barriers at public vocational schools highlight the importance of external factors. A hypothetical model addresses this within the pillars of strengthening resources and meaningfulness. Research by Sudira (2018) on Vocational Education and Training (VET) in various countries shows that successful VET systems are characterized by close business involvement in the assessment process. The presence of industry assessors not only enhances the validity and credibility of certificates but also serves as a powerful motivator for students, as the model proposes. The vocational education assessment-strengthening model developed in this study represents a systematic approach to addressing the multidimensional challenges in assessing vocational high school students' competencies. Philosophically, this model aligns with the concept of assessment for learning (Ahin et al., 2023), which emphasizes the formative role of assessment as a catalyst for learning improvement. The conceptual transformation from "assessment" to "competency testing" or "mini certification" creates a sense of urgency and relevance to the world of work, as evidenced by Tsai's (2024) research on the effectiveness of industrial simulation approaches. Within the framework of strengthening the assessment process, implementing an integrative, continuous cycle aligns with the principles of sustainable assessment. The proposed Integrated Practical Exam innovation is a manifestation of integrated assessment, capable of measuring competency holistically through authentic tasks. This approach assesses not only the final product but also the work process and the development of professional attitudes, in line with the demands of 21st-century competencies.

The strategy of strengthening resources through optimization and collaboration reflects the findings of Pambayun et al. (2023) on the

importance of strategic resource management in vocational education. The establishment of an industrial workshop simulation with equipment arranged by work area aligns with the recommendations of Prianto et al. (2021). Strengthening teacher capacity through specialized training and the development of structured instruments reinforces that the primary determinant of assessment quality lies in teachers' pedagogical competence. The development of an attitude assessment rubric with clear indicators is a response to the often-subjective practice of affective assessment. The integration of structured peer assessment aligns with Jiang et al. (2023) findings that peer assessment can enhance students' metacognition and responsibility for learning.

In other words, this model integrates critical elements necessary for comprehensive vocational assessment reform and aligns with the comprehensive assessment reform framework (Nadzar & Supriyadi, 2020). Implementation of this model is expected to address the gap between conventional assessment practices and industry demands while comprehensively developing both hard and employability skills. Successful implementation of this model requires strong institutional commitment and sustained policy support, as emphasized in the context of meaningful education reform.

■ CONCLUSION

Based on the findings, assessment problems in the cognitive, psychomotor, and affective domains in the two vocational high schools studied are interrelated. In the cognitive domain, teachers faced challenges related to low student literacy engagement, limited assessment methods dominated by multiple-choice tests, and the suboptimal use of diagnostic assessment to improve learning. In the psychomotor domain, assessment was constrained by limited facilities, limited time, equipment wear, and uneven

opportunities to observe each student's performance individually. In the affective domain, teachers had difficulty conducting in-depth attitude assessments due to large class sizes, limited observation time, and the lack of explicit, structured instruments for attitude assessment. These findings are important for vocational education because they show that assessment quality is closely related to student competence, workplace readiness, and the alignment between school learning and industry expectations.

The study also shows an imbalance between formative and summative assessment functions. Formative assessment is often sacrificed to meet material completion targets and school agendas, so it is not fully utilized as a mechanism for feedback and instructional improvement. Summative assessment is burdened by administrative demands and pressure to produce final grades, while remedial programs tend to be limited to retake exams without sufficient additional learning. In response to these issues, this study proposes a hypothetical model to strengthen vocational learning assessment through four pillars: enhancing the meaningfulness of assessment, strengthening the diagnostic-formative-summative assessment cycle, strengthening resources through collaboration, and strengthening teacher and instrument capacity.

This study implies that vocational schools need to redesign assessment as a competence-building process rather than merely a grading activity. Schools should support teachers through practical assessment training, structured rubrics, scheduled competency test days, team teaching, item analysis follow-up, and stronger industry involvement. However, this study is limited to two schools and six teachers in the Automotive Engineering Expertise Program, so the findings should be interpreted as contextual rather than generalized to all vocational high schools. The proposed model remains hypothetical, although it has been discussed in a subject teacher forum

(MGMP). Further research is needed to validate the model and test its effectiveness with a broader range of participants, across different vocational fields, and using mixed-methods or development research designs.

■ DECLARATION OF GENERATIVE AI USAGE IN THE WRITING PROCESS

Declaration of Generative AI Usage in the Writing Process

During the writing of this manuscript, the author(s) employed Gemini for a specific purpose, e.g., language refinement/proofreading. The author(s) have reviewed and edited the content generated by this tool and assume full responsibility for the published article.

■ REFERENCES

- Abdurrahman, I. S., & Mahmudah, F. N. (2023). Development of a digital-preneurship measurement Instrument: Alignment approach through project-based learning. *International Journal of Educational Methodology*, 9(1), 283–295. <https://doi.org/10.12973/IJEM.9.1.283>
- Airasian, P. W. (2001). *Classroom assessment: Concepts and applications* (4th ed.). McGraw-Hill. <https://eric.ed.gov/?id=ED446097>
- Ayub, M. R. S. S. N., Istiyono, E., Munadi, S., Permadi, C., Pattiserlihun, A., & Sudjito, D. N. (2020). *Analisa penilaian soal fisika menggunakan model Rasch dengan program R* [Analysis of physics question assessment using the Rasch model with the R program]. *Jurnal Sains dan Edukasi Sains*, 3(2), 46–52. <https://doi.org/10.24246/juses.v3i2p46-52>
- Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & van der Vleuten, C. P. M. (2007). Evaluating assessment quality in competence-based education: A qualitative

- comparison of two frameworks. *Educational Research Review*, 2(2), 114–129. Doi:10.1016/j.edurev.2007.6.001
- Baartman, L. K. J., Prins, F. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2007). Determining the quality of competence assessment programs: A self-evaluation procedure. *Studies in Educational Evaluation*, 33 (3-4), 258–281. <https://doi.org/10.1016/j.stueduc.2007.07.004>
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80 (2). <https://doi.org/10.1177/003172171009200119>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Chappuis, S., & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership*. 60(1). <http://hssdnewteachers.pbworks.com/w/file/fetch/50394085/classroom.assessment.for.learning.chappuis.pdf>.
- Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57–66. <https://doi.org/10.1080/14703290601081332>
- Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches (2nd ed.)*. Sage Publications, Inc.
- Firman, N., Sudira, P., & Minghat, A.D. (2018). Literature review: Authentic assessment in vocational high schools. *International Journal of Engineering & Technology*. 7(4): 191–194. DOI:10.14419/ijet.v7i4.33.23557
- Fullan, M., & Quinn, J. (2015). *Coherence: The right drivers in action for schools, districts, and systems*. Corwin Press.
- Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52(3), 67–86. doi:10.1007/BF02504676
- Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Corwin Press. <https://doi.org/10.4135/9781452219493>
- Istiyono, E. (2020). *Pengembangan instrumen penilaian dan analisis hasil belajar fisika dengan teori tes klasik dan modern (Edisi Kedua)* [Development of assessment instruments and analysis of physics learning outcomes using classical and modern test theories (Second Edition)]. UNY Press.
- Jiang, C., Ma, H., Luo, Y., Fong, D. Y. T., Umucu, E., Zheng, H., Zhang, Q., Liu, X., Liu, X., Spruyt, K., & Meng, R. (2023). Validation of the Chinese version of the Perceived Stress Scale-10 integrating exploratory graph analysis and confirmatory factor analysis. *General Hospital Psychiatry*, 84, 194–202. <https://doi.org/10.1016/j.genhosppsych.2023.07.008>
- Lin, P. Y., Chen, P. Y., Liu, Y. C., & Yang, H. C. (2024). The impact of the Ok4R mobile platform on vocational students' improvement of science reading comprehension and logical thinking. *Journal of Baltic Science Education*, 23(1), 90–103. <https://doi.org/10.33225/jbse/24.23.90>
- Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan non tes* [Techniques for developing test and non-test instruments]. Mitra Cendikia Offset.
- Mardapi, D. (2017). *Pengukuran, penilaian*

- dan evaluasi pendidikan (Edisi ke-2) [Measurement, assessment, and evaluation of education (2nd ed.)]. Parama Publishing.
- McMillan, J. H. (2018). *Classroom assessment: Principles and practice that enhance student learning and motivation* (7th ed.). Pearson.
- Misbah, Z., Gulikers, J., Dharma, S., & Mulder, M. (2020). Evaluating competence-based vocational education in Indonesia. *Journal of Vocational Education & Training*, 72(4), 488–515. <https://doi.org/10.1080/13636820.2019.1635634>
- Mustapha, R., Sadrina, Nashir, I. M., Azman, M. N. A., & Hasnan, K. A. (2020). Assessing the implementation of project-based learning (PjBL) in the Department of Mechanical Engineering at a Malaysian Polytechnic. *Journal of Technical Education and Training*, 12(1 Special Issue), 100–118. <https://doi.org/10.30880/jtet.2020.12.01.011>
- Nadzar, A. H., & Supriyadi, E. (2020). *Pengembangan media pembelajaran berbasis android database komponen listrik* [development of android-based learning media for electrical component databases]. *Jurnal Edukasi Elektro*, 4(2), 103–109. <https://doi.org/10.21831/jee.v4i2.35855>
- Nisa, C., & Retnawati, H. (2018). Comparing the methods of vertical equating for the math learning achievement tests for junior high school students. *REID (Research and Evaluation in Education)*, 4(2), 164–174. <https://doi.org/10.21831/reid.v4i2.19291>
- Ordóñez-Ávila, R., Salgado Reyes, N., Meza, J., & Ventura, S. (2023). Data mining techniques for predicting teacher evaluation in higher education: A systematic literature review. In *Heliyon* (Vol. 9, Issue 3). Elsevier Ltd. <https://doi.org/10.1016/j.heliyon.2023.e13939>
- Pals, F. F. B., Tolboom, J. L. J., & Suhre, C. J. M. (2023). Development of a formative assessment instrument to determine students' need for corrective actions in physics: Identifying students' functional level of understanding. *Thinking Skills and Creativity*, 50. <https://doi.org/10.1016/j.tsc.2023.101387>
- Pambayun, N. A. Y., Munadi, S., Arifin, Z., Setiawan, C., & Retnawati, H. (2023). Industrial work culture education in Indonesian vocational high schools: Teachers' perceptions and practices. *Issues in Educational Research*, 33(2), 713–732. <https://search.informit.org/doi/abs/10.3316/informit.303109822674581>
- Pratama, R. A., Pratiwi, I. M., Saputra, M. A., & Sumargono. (2022). Integration of STEM education in history learning. *International Journal of Evaluation and Research in Education*, 11(1), 313–320. <https://doi.org/10.11591/ijere.v11i1.22064>
- Prianto, A., Winardi, W., & Qomariyah, U. N. (2021). The effect of the implementation of teaching factory and its learning involvement toward work readiness of vocational school graduates. *International Journal of Instruction*, 14(1): 283–302. DOI: 10.29333/iji.2021.14117a
- Presidential Regulation No. 8 of 2012 concerning the Indonesian National Qualifications Framework.
- Randhawa, J. S., & Ahuja, I. S. (2017). 5S - A quality improvement tool for sustainable performance: Literature review and directions. *International Journal of Quality & Reliability Management*, 34(3), 334–361. <https://doi.org/10.1108/IJQRM-03-2015-0045>
- Putra, B. J. (2023). Study of stake evaluation

- model in the interest of program evaluation. *Jurnal Sosial Humaniora Sigli*. 6(1), 220–223. <https://doi.org/10.47647/jsh.v6i1.1489>
- Rezkilaturahmi, R., Danni, R., & Retnawati, H. (2024). Investigating test equating methods: English examination test. *International Journal of Education and Curriculum Application* 7(3), 246–257. DOI:10.31764/ijeca.v7i3.25065
- Rezkilaturahmi, R., Istiqlal, M., Putro, N.H.P., & Istiyono, E. (2025). Evaluating English language test items developed by teachers: An item response theory approach. *VELES Voice of English Language Education Society*, 9(1), 218–230. DOI:10.29408/veles.v9i1.27644
- Santoso, A., Retnawati, H., Pardede, T., Apino, E., Rafi, I., Rosyada, M., Kassymova, G. & Wenxin, X. (2024). From investigating the alignment of a priori item characteristics based on the CTT and four-parameter logistic (4-PL) IRT models to further exploring the comparability of the two models. *Practical Assessment, Research, and Evaluation*. 29(1). <https://doi.org/10.7275/pare.2043>
- Retnawati, H., Kardanova, E., Sumaryanto, Prasojo, L. D., Jailani, Arliani, E., Hidayati, K., Susanti, M., Lestari, H. P., Apino, E., Rafi, I., Rosyada, M. N., Tuanaya, R., Dewanti, S. R., Sotlikova, R., & Kassymova, G. K. (2024). A Systematic review of the use of technology in educational assessment practices: Lesson learned and direction for future studies. *International Journal of Robotics and Control Systems*, 4(4), 1656–1693. <https://doi.org/10.31763/ijrcs.v4i3.1572>
- Pahin, M. G., Yýldýrym, Y., & Öztürk, N. B. (2023). Examining the achievement test development process in the educational studies. *Participatory Educational Research*, 10(1), 251–274. <https://doi.org/10.17275/per.23.14.10.1>
- Setiawan, H. J., & Islami, N. (2020). Improving critical thinking skills of senior high school students using the problem based learning model. *Journal of Physics: Conference Series*, 1655(1). <https://doi.org/10.1088/1742-6596/1655/1/012060>
- Sudira, P. (2018). *Metodologi pembelajaran vokasional abad XXI: Inovasi, teori dan praksis. [21st Century vocational learning methodology: Innovation, theory and practice]*. UNY Press.
- Suhartanta, S., Efendi, Y., & Ramadhan, A. (2023, March). Needs assessment of competency for human resource development in vehicle wheel alignment workshop. In *AIP Conference Proceedings* (Vol. 2671, No. 1, p. 020025). AIP Publishing LLC. <https://doi.org/10.1063/5.0115997>
- Thanabalan, T. V., Siraj, S., & Alias, N. (2015). Evaluation of a digital story pedagogical module for the indigenous learners using the stake countenance model. *Procedia-Social and Behavioral Sciences*, 176, 907–914.
- Thorndike, E.L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1), 25–29. <https://doi.org/10.1037/h0071663>
- Tosuncuoglu, I. (2018). Importance of assessment in ELT. *Journal of Education and Training Studies*, 6(9), 163. <https://doi.org/10.11114/jets.v6i9.3443>
- Tsai, F.-H. (2024). Development and evaluation of an Internet of Things project for preservice elementary school teachers. *Sustainability*, 16(17), 7632. <https://doi.org/10.3390/su16177632>
- Wagner, F. L., Sudacka, M., Kononowicz, A. A., Elvén, M., Durning, S. J., Hege, I., & Huwendiek, S. (2024). Current status and

- ongoing needs for the teaching and assessment of clinical reasoning – an international mixed-methods study from the students' and teachers' perspective. *BMC Medical Education*, 24(1), 1–17. <https://doi.org/10.1186/s12909-024-05518-8>
- Wiliam, D. (2011). What is assessment for learning? *Studies in educational evaluation*, 37(1), 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- Zhang, D., & Hwang, G. J. (2023). Effects of interaction between peer assessment and problem-solving tendencies on students' learning achievements and collaboration in mobile technology-supported Project-Based Learning. *Journal of Educational Computing Research*, 61(1), 208–234. <https://doi.org/10.1177/07356331221094250>