Jurnal Penelitian dan Pengkajian Ilmu Pendidikan: e-Saintika
https://journal-center.litpam.com/index.php/e-Saintika/index

November 2026 Vol. 10, No. 3
e-ISSN: 2615-6881
pp. 1127-1158

Trainer-Based Engine Management System Diagnostic Learning Package Development and Evaluation in Vocational High Schools

*** Robi Irza Saputra, Rifdarmon, Arwizet Karudin, Toto Sugiarto**

Technology and Vocational Education Study Program, Faculty of Engineering, Universitas Negeri Padang. Jl. Prof. Dr. Hamka, Air Tawar Barat, Padang 25131, Indonesia.

* Corresponding Author e-mail: robisaputra22@guru.smk.belajar.id

Received: August 2026; Revised: November 2026; Published: November 2026

Abstract

Electronic engine diagnosis requires vocational students to interpret sensor–ECU–actuator interactions and verify faults using diagnostic evidence, but hands-on measurement opportunities are limited. This study developed and evaluated a structured diagnostic learning package comprising a practicum guide, worksheets, fault scenarios, lesson sequences, and assessment instruments for a commercially available Engine Management System (EMS) trainer. The 4D development model was combined with a nonequivalent pretest–posttest design involving two intact Grade XI classes at one Indonesian vocational high school (35 students per class). Eight experts rated the package very valid (91.33%–92.79%), and all ten inspection-based function and safety criteria were met. Teacher (n = 3) and student (n = 15) practicality ratings were favourable (pooled 90.97%), although the student questionnaire showed poor internal consistency (Cronbach's $\alpha = -0.41$). Cognitive scores improved in both classes (Wilcoxon $p < 0.001$), but baseline achievement differed ($p = 0.042$), and between-class raw-gain ($p = 0.102$) and normalized-gain ($p = 0.074$) comparisons were not significant. The additive pretest-adjusted estimate was inconclusive (0.82 points; 95% CI −0.25 to 1.90), with a significant class-by-pretest interaction limiting its interpretation. In a group-based assessment of one practised ECT fault, experimental students achieved a mean observed troubleshooting score of 88.60. Technical checks documented DTC responses without substantial voltage changes at several ECU test points, and the ECT assessment reading was not reproduced. The study contributes a task-aligned learning package and distinguishes trainer functions, local feasibility, and learning outcomes; causal superiority, independent diagnostic competence, and transfer to unfamiliar faults remain unestablished.

Keywords: Diagnostic Reasoning; Engine Management System; Fault Simulation; Hardware Trainer; Vocational Education

How to Cite: Saputra, R. I., Rifdarmon, R., Karudin, A., & Sugiarto, T. (2026). Trainer-Based Engine Management System Diagnostic Learning Package Development and Evaluation in Vocational High Schools. *Jurnal Penelitian Dan Pengkajian Ilmu Pendidikan: E-Saintika, 10*(3), 1127-1158. https://doi.org/10.36312/e-saintika.v10i3.7123

10.36312/e-saintika.v10i3.7123

Copyright© 2026, Saputra et al.
This is an open-access article under the CC-BY-SA License.

INTRODUCTION

Vocational education is expected to develop technical competence that transfers to the workplace, so learning experiences should approximate authentic work situations rather than stop at conceptual knowledge (Gupta et al., 2024). In automotive programs, the spread of electronic engine control has shifted the core of workshop competence from mechanical adjustment towards electrical measurement and fault diagnosis (Purnama & Wagiran, 2020). In Indonesian vocational high schools (*Sekolah Menengah Kejuruan*, SMK), this competence is taught in the Light Vehicle Engineering program (*Teknik Kendaraan Ringan*, TKR) under the Engine System element, in which the Engine Management System (EMS) is a central topic: sensors convert engine

conditions into electrical signals, an electronic control unit (ECU) processes them, and actuators execute the ECU's commands.

Diagnosing faults in such a system is a specific and demanding form of problem solving. Studies of automotive apprentices and technicians show that successful diagnosis depends on a usable model of the system, on systematic collection of evidence, and on testing hypotheses before replacing parts, whereas less successful diagnosticians tend to jump from a symptom or a diagnostic trouble code (DTC) to component replacement (Abele, 2018; Schaafstal et al., 2000). Troubleshooting is therefore learned best when learners repeatedly move between a system model, measurements, and hypotheses about the location of a fault (Jonassen & Hung, 2006). In the EMS topic, three conditions make this difficult for students: the signal flow between sensors, ECU, and actuators cannot be observed directly; on a real vehicle, sensors, connectors, and harnesses are hidden in the engine bay and are hard to measure without disassembly; and a DTC only identifies a circuit to be checked, so students must learn to confirm it with voltage, resistance, and continuity measurements.

A needs analysis at the research school confirmed these conditions. Three teachers of the Engine System element (questionnaire, 22 January–4 February 2026) gave the lowest score (1.00 on a five-point scale) to the availability of interactive EMS media, the ownership of a trainer, an adequate trainer–student ratio, and media that visualize the sensor–ECU–actuator flow. Students (n = 64; 16 and 24 February 2026) rated their understanding of EMS under current teaching at 1.58, reported great difficulty imagining sensor and ECU operation from verbal explanation (4.86), and expressed a strong need for direct measurement practice (5.00). These questionnaires constituted the needs analysis; the proportions of students scoring below 75 that were reported in the previous version of this manuscript (48.57% and 31.43%) come from the pretest administered in July 2026 and are now reported only in the Results section to keep the chronology clear.

Research on media for automotive diagnosis has followed two lines (Table 1). Digital media for EMS, such as 4D-model multimedia (Fitriyanto et al., 2020) and interactive scan-tool multimedia (Mubasir, 2020), improved test scores but mediate the experience through representations, so students do not take real measurements. Physical media include an engine simulator trainer used with problem-based learning (Tohadi et al., 2020), a Toyota K3-VE EMS training kit evaluated on 20 students with an N-Gain of 0.62 (Handoyono & Rabiman, 2023), a diagnostic trouble box that improved analytical thinking in a quasi-experiment with 86 automotive students (Perdana Sutrisno et al., 2025), and trainers for neighbouring electrical subsystems such as lighting, charging, and electric drive (Alfana et al., 2024; Firmansyah & Arizal, 2026; Rohayati et al., 2025; Satriawan et al., 2025; Siswanto et al., 2024; Wawan & Arizal, 2026). A simulator-based CVT study (Taufiq et al., 2024) and a problem-based teaching-factory model for automotive problem solving (Maksum et al., 2024) show the same pattern. Most of these studies report expert feasibility and test-score gains; few describe the technical behaviour of the device, test whether groups were equivalent at baseline, or assess the diagnostic process itself. Computer-based assessments of repair and diagnostic performance have also shown that scores obtained in simplified settings do not automatically predict hands-on performance (Hartmann et al., 2023), which underlines the need to describe how closely a trainer reproduces the diagnostic task.

Recent studies further illustrate the range of approaches to automotive learning media. Hidayat et al. (2024) developed a QR-supported electronic fuel injection (EFI) simulator and reported favourable expert assessment and functional observations. Nanda et al. (2025) integrated a mobile QR-EFI simulator with a problem-based teaching-factory model and reported improvement in 4C skills. Liao et al. (2023) investigated virtual-reality-supported vehicle electrical maintenance using a zone-of-proximal-development instructional strategy. Although these studies support ongoing developments in automotive education, their systems, outcomes, and instructional arrangements differ from the terminal-level EMS diagnostic learning package examined in this study.

Table 1. Comparison of previous automotive diagnostic media with the learning package developed in this study

Study	Medium and system	Diagnostic learning activities reported	Evidence reported	Main limitation
Fitriyanto et al. (2020); Mubasir (2020)	Digital multimedia; EMS, scan tool	Viewing simulated measurements and scan-tool screens; no physical fault localization or verification	Validity; test gain	No physical measurement
Tohadi et al. (2020)	Engine simulator trainer (4E-FTE)	Problem-based cases on a running engine; measurement-based fault localization and verification not described	Test gain	Signal measurement not documented
Handoyono & Rabiman (2023)	EMS training kit, Toyota K3-VE	Practicum with real EMS components; worksheets for fault localization and verification not described	Feasibility; N-Gain 0.62 (n = 20)	No control group
Perdana Sutrisno et al. (2025)	Diagnostic trouble box	Simulation-based diagnosis of fault scenarios; EMS signal measurement at ECU terminals not described	Analytical thinking vs. control	Not an EMS signal path
Alfana et al. (2024); Rohayati et al. (2025); Wawan & Arizal (2026)	Lighting and charging simulators	Practicum on circuit faults in other electrical subsystems	Validity; test gain or mastery	Other subsystems
Satriawan et al. (2025); Firmansyah & Arizal (2026)	Electric-vehicle kits/trainers	Component demonstration and practicum	Feasibility; mastery	Other subsystem
This study	Commercial EMS panel integrated with a structured diagnostic learning package	Worksheet-guided identification and available terminal measurements; DTC retrieval and interpretation; fault investigation and restoration; collaborative	Expert validity; function and safety checks; descriptive practicality; cognitive outcomes; observed group-	One school/class per condition; baseline imbalance; uneven voltage responses; one practised group-assessed fault

Study	Medium and system	Diagnostic learning activities reported	Evidence reported	Main limitation
		verification on a practised scenario	based troubleshooting	

Comparisons of physical, virtual, and remote laboratories show that hands-on equipment is not superior by default; its value depends on what learners actually do with it, especially whether they collect and interpret real data (Brinson, 2015; de Jong et al., 2013; Ma & Nickerson, 2006; Zacharia & Olympiou, 2011). Laboratory objectives in engineering similarly emphasize instrumentation, measurement, and learning from failure (Feisel & Rosa, 2005). The simulation literature adds that functional fidelity, the degree to which a training device reproduces the cues and responses needed for the task, matters more than physical resemblance (Chernikova et al., 2020; Hamstra et al., 2014). For EMS diagnosis, the functionally relevant cues are reference and signal voltages, ground and supply conditions, DTCs, and the response of actuators to the ECU.

Technical evidence also warrants closer attention. Fault-injection research has examined how failures produce observable system responses: Lee et al. (2022) analysed fault-mode–effect relationships in an electronic fuel-injection model, and Ghimire et al. (2018) used fault-injection experiments to link disturbances with sensor measurements in an electric power-steering system. Although these are engineering rather than education studies, they establish a rationale for distinguishing the presence of a diagnostic code from independently measured evidence of a particular electrical fault.

Two related research gaps emerge from the reviewed literature. First, learning-media evaluations seldom document in comparable detail which diagnostic indications and electrical measurements are actually reproduced at accessible points of a commercial EMS trainer, or how those functions are integrated into structured instructional tasks. Second, validity, favourable usability ratings, cognitive test gains, and observed troubleshooting performance are often discussed without sufficiently separating what each measures. This distinction is crucial when assessing intact classes with different initial achievement and practical tasks completed collaboratively. The present investigation addresses these gaps at a single vocational school; it does not claim that no related approaches exist elsewhere.

The study is framed by experiential learning, in which competence develops through cycles of concrete experience, reflective observation, conceptualization, and active experimentation (Kolb & Kolb, 2005; Morris, 2020), and by research on authentic assessment, which requires assessment tasks to resemble the professional task (Gulikers et al., 2004). Figure 1 shows the resulting conceptual model. Trainer characteristics (real components, visible wiring, test points, fault switches, and OBD-II access) enable learning activities (observing, measuring, reading DTCs, hypothesising, and verifying); these activities are expected to work through concrete experience, reflection on measured data against reference values, and structured hypothesis testing (Schaafstal et al., 2000); and they are assessed as cognitive understanding and observed troubleshooting performance. Validity and practicality are treated as evidence about the product, not about learning.

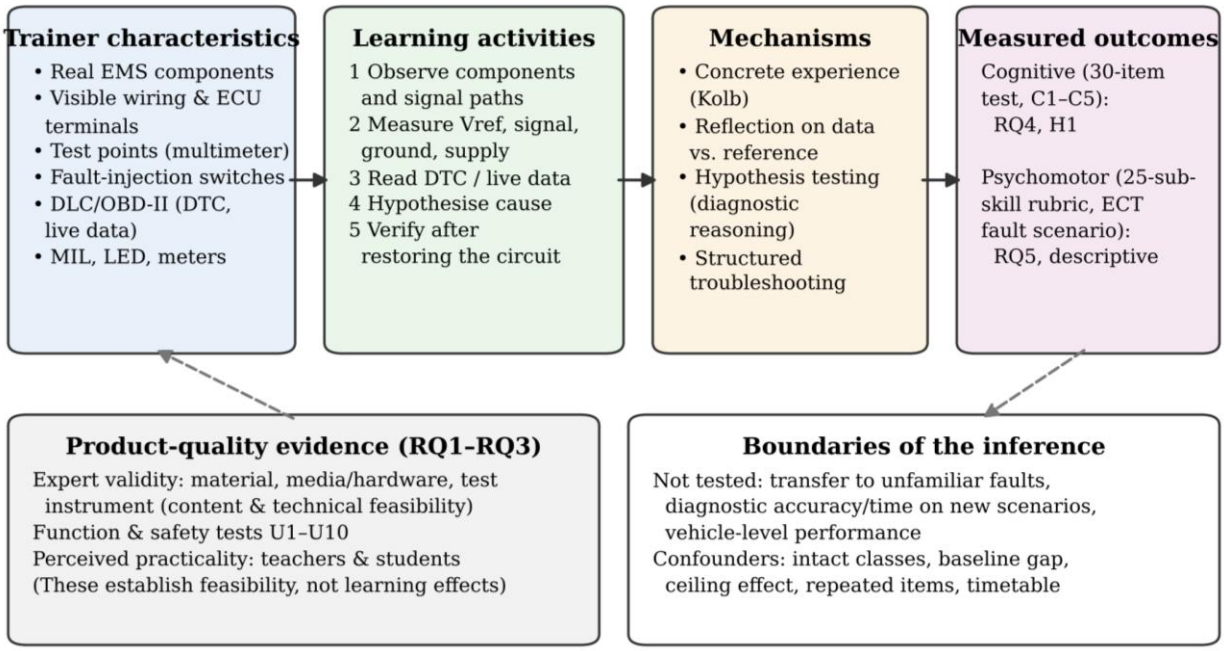


Figure 1. Conceptual model of the study

The study contributes an instructional design and transparent evaluation of a commercially available EMS trainer, not a newly designed hardware platform. Its learning package includes a practicum guide, worksheets, graded fault scenarios, lesson sequences, and assessment instruments. The first contribution is to translate structured diagnostic reasoning into activities involving measurement, diagnostic feedback, fault hypotheses, and verification. The second is to characterize the difference between DTC generation and electrically observable fault effects. The third is to evaluate expert validity, user-perceived practicality, cognitive learning, and observed troubleshooting as distinct evidence. The inspection checklist adapts existing technical criteria rather than introducing a new theoretical instrument.

In the instructional sequence, Worksheet 1 guides students to identify sensor terminals and compare available reference, signal, and ground measurements with documented values. Worksheet 2 links manually varied TPS input to actuator responses and asks students to interpret ECT-related indications; because the panel does not dynamically heat coolant, ECT activities do not simulate a real temperature transition. Worksheet 3 addresses DTC and live-data interpretation and requires students to justify suspected causes using the available circuit evidence. The final diagnostic scenario follows complaint, plan, DTC, measurement, cause analysis, restoration, and verification. This application of diagnostic reasoning draws on Jonassen (2000) and Schaafstal et al. (2000), while recognizing that not every fault switch produced a measurable terminal-voltage change.

The study addressed five research questions: (RQ1) How can a structured diagnostic learning package for a commercial EMS trainer be developed under the 4D model, and which technical functions and limitations are documented? (RQ2) How valid is the package according to material, media/hardware, and learning/test experts? (RQ3) How do teachers and students rate its practicality? (RQ4) How do observed cognitive changes compare between the trainer-supported and conventional classes, accounting for baseline differences? (RQ5) What troubleshooting performance is observed in a practised scenario? For RQ4, H1 specified a difference in observed

improvement between classes (H0: no difference); this was not a causal treatment-effect hypothesis. RQ5 was descriptive.

METHOD

Research Design and Development Process

This study employed the Define, Design, Develop, and Disseminate (4D) research-and-development model of Thiagarajan et al. (1974). The model was applied to a diagnostic learning package, not to the fabrication of hardware. During Define, a needs analysis specified required features: actual EMS components, accessible terminal test points, fault-switch simulation, and OBD-II diagnosis. The school then purchased a commercially available trainer matching those requirements. During Design and Develop, the researchers produced, validated, and revised the practicum guide, worksheets, fault scenarios, lesson plans, and assessment instruments. Hardware-related work consisted of functional inspection and documentation of the purchased unit. The Disseminate stage was limited to classroom use at the research school rather than wider diffusion.

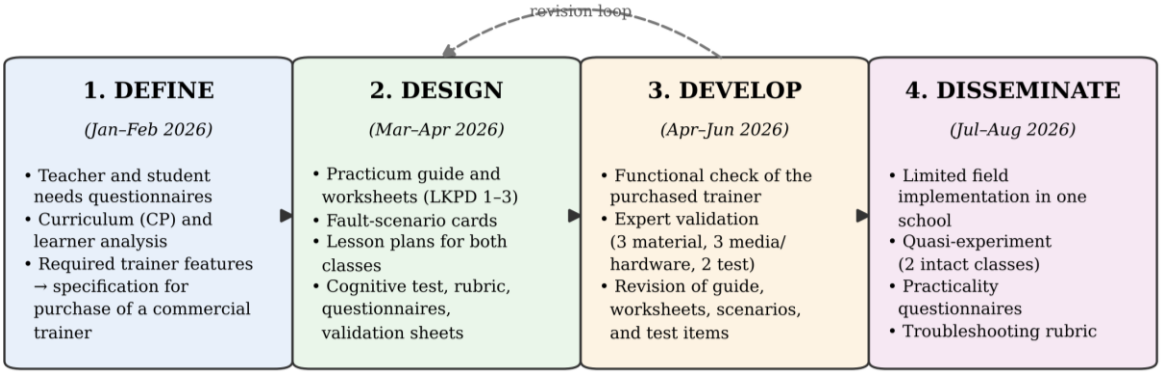


Figure 2. The 4D model as applied in this study.

The classroom assessment used a quasi-experimental nonequivalent control-group pretest–posttest design because students belonged to intact classes that could not be randomly reorganized (Kim & Steiner, 2016). Treatment was assigned to one class per condition. Therefore, treatment and class membership were completely confounded; student-level comparisons describe outcomes in the two observed classes but cannot identify a class-independent causal treatment effect. Figure 2 shows the development stages and their timing, and Table 2 integrates the class allocation, teaching sequence, dates, and assessments.

Table 2. Quasi-experimental design, instructional activities, and assessment schedule for the experimental and control classes

Meeting/date	Learning focus	Experimental: XI TKR 2 (n = 35)	Control: XI TKR 1 (n = 35)	Assessment
Meeting 1 Exp: 16 Jul 2026 Ctrl: 13 Jul 2026	EMS concepts and sensors (TP 1–2)	Pretest; safety briefing; component and wiring identification; available test-point	Pretest; lecture, diagrams and slides; worked ECT example; written exercises	Pretest (O1); Worksheet 1

Meeting/date	Learning focus	Experimental: XI TKR 2 (n = 35)	Control: XI TKR 1 (n = 35)	Assessment
		measurements for MAP, IAT, ECT, TPS and CKP/CMP		
Meeting 2 Exp: 23 Jul Ctrl: 20 Jul	Actuators and sensor–ECU–actuator flow (TP 3)	Manual variation of TPS input; interpretation of ECT-related diagnostic cues; observation of actuator/indicator response; flow diagrams	Lecture on injectors, coils, ISC and ground switching; paper flow diagrams and exercises	Worksheet 2; five-item formative quiz (not a research outcome)
Meeting 3 Exp: 30 Jul Ctrl: 27 Jul	DTC, MIL, DLC/OBD-II and live data (TP 4)	Scan-tool connection; guided fault simulation (2–3 cases/group); comparison of DTC/live data and available measurements	DTC and live-data lecture; discussion of written cases	Worksheet 3; scanner performance check (experimental)
Meeting 4 Exp: 6 Aug Ctrl: 3 Aug	Basic troubleshooting (TP 5)	Group-based customer complaint: symptom → DTC → measurement → cause → verification; reflection	Troubleshooting lecture, written fault cases, worksheet and reflection	Troubleshooting rubric (experimental only)
Follow-up / day 28 Exp: 13 Aug Ctrl: 10 Aug	Cognitive outcomes	Posttest and practicality questionnaire (15 students)	Posttest	Posttest (O2); student practicality

Research Setting, Participants, and Group Allocation

The primary study, including instructional development, needs analysis, expert validation, and classroom data collection, was conducted at SMK Negeri 1 Bangkinang, Riau, Indonesia, between January and August 2026. A separate post-implementation technical inspection was conducted after the classroom data collection had ended, as described below. Three Engine System teachers completed needs-analysis questionnaires between 22 January and 4 February 2026. Student needs-analysis questionnaires were completed by 64 students (30 from XI TKR 1 on 16 February and 34 from XI TKR 2 on 24 February); the remaining students in the two classes did not respond. This sample is distinct from the 70 students subsequently assessed for cognitive outcomes. Expert validation involved three university material experts, three media/hardware specialists (two dealership trainers/coordinators and

one workshop lead mechanic), and two learning/test experts who were vocational teachers. Practicality was assessed by three teachers and 15 experimental-class students who were also included in the effectiveness sample.

The two Grade XI Light Vehicle Engineering classes were selected purposively and retained their original membership (35 students per class). XI TKR 2 was assigned to trainer-supported instruction, and XI TKR 1 to conventional instruction, before the pretest and without random allocation or reference to prior achievement. Both classes were taught by the same instructor, who also coordinated the classroom implementation and assessment. The control timetable was Monday 08:00–11:10, and the experimental timetable Thursday 13:10–15:50. Although the instructional plans allocated four 40-minute periods to each session, actual uninterrupted learning time was not logged. Differences in baseline performance, timetable and class characteristics could influence comparisons; with only one class per condition they could not be distinguished from intervention effects.

Trainer Specification and Technical Characterization

The school purchased a commercial vertical EMS trainer panel mounted on a mobile steel frame (Figure 3). The researchers specified the desired instructional features, inspected the unit, and documented selected functions; they did not design or fabricate the hardware. Table 3 identifies its sensors, ECU terminals, fault switches, output devices, and operating conditions. The panel uses a 12 V/32 Ah battery charged by a 220 V AC supply and is protected by fuses and relays. The Toyota Avanza K3 ECU communicates through a G-scan 3 using Diagnosis → Daihatsu → International → Diagnosis → DTC Analysis.

With ignition ON, the MIL illuminates during the bulb check. At START, an electric motor drives the belt-coupled CKP/CMP reluctor assembly; the ECU controls ignition coils and injectors, producing visible sparks and spray into transparent tubes. The speed regulator changes the displayed rotational speed. These functions simulate selected electrical and diagnostic interactions, not combustion, load, changing coolant or air temperature, intake vacuum, or closed-loop fueling. Oxygen and vehicle speed sensors are not fitted to the panel. Figure 4 shows terminal sockets and fault switches; Figure 5 distinguishes functional pathways from the internal wiring, which was not independently traced.

According to the trainer manual, moving a designated fault switch physically to OFF activates the simulated fault and can generate the documented DTC; restoring ON and using Erase clears the indication. DTCs do not independently identify a specific short or open circuit. The manual does not provide the internal fault-switch schematic, so fault mechanisms were not assumed from a generic code. The validation form's wording 'fault simulation ON' describes an active fault, not the physical switch position. Electrical readings were taken with a digital multimeter between the relevant labelled ECU banana-socket terminal and panel ground, with ignition ON.

Table 3. Technical specification of the EMS trainer

Component (Fig. 3b no.)	Type / operating principle	ECU terminal	Fault switch → DTC	Input generation on trainer
ECT sensor (1)	NTC thermistor, 5 V reference	THW	ECT → P0115	Real sensor at ambient panel temperature (no heating); Vref 5 V ± 0.2

Component (Fig. 3b no.)	Type / operating principle	ECU terminal	Fault switch → DTC	Input generation on trainer
				V; 2–3 kΩ at room temperature
IAT sensor (2)	NTC thermistor, 5 V reference	THA	IAT → P0110	Real sensor at ambient temperature
MAP sensor (3)	Pressure-sensing element, 5 V reference	PIM, VCPM	MAP → P0105	No vacuum source on panel; static (atmospheric) input
Throttle body with TPS and ISC valve (4)	Potentiometer (TPS); rotary idle-air valve (ISC)	VTH, VC; IACA/IACB	TP → P0120; ISC → P0505	Throttle opened by hand; ISC driven by ECU
Knock sensor (5)	Piezoelectric element	KNK	KNK (no DTC)	No knock occurs, so no signal or DTC is produced
Oil control valve (6)	ECU-driven solenoid	OCV+, OCV−	OCV → P0010	Driven by ECU
CKP and CMP sensors (8)	Pick-up sensors on a belt-coupled reluctor assembly	N1+/N1−, N2+/N2−	–	Electric drive motor with speed regulator; speed shown on tachometer
ECU (14) and terminal test points (15)	Toyota Avanza K3 engine ECU; banana-socket test point on each signal line	All listed	–	Scan-tool protocol: Daihatsu (K3 engine)
Injectors #1–#4 (13)	Ground-switched solenoids; fuel rail with transparent spray tubes; electric fuel pump	#10–#40	–	Supply 12 V with ignition ON; spray visible when cranking
Ignition coils #1–#4 (12)	Coil-on-plug, ECU-switched primary	IG1–IG4	–	Spark visible at plugs when cranking
DLC (16), MIL (10)	OBD-II data link connector; MIL	–	–	G-scan 3 reads and erases DTCs
Power and protection (17, 19)	Battery 12 V/32 Ah charged from 220 V AC; AM/IG1/IG2/+B circuits; 6 fuses; 3 relays; ignition switch	–	–	–
Instruments (9, 11, 18)	Tachometer, voltmeter, ammeter, fuel-pressure gauge	–	–	Displayed on panel

During the study, students recorded a reference/pull-up voltage of 4.87–5.12 V at THW with the ECT sensor connector disconnected and the ignition ON. In the assessed fault condition, they recorded 0–0.12 V at THW with the connector fitted and ECT switch OFF; these measurements involved different connector conditions. A supplementary technical inspection on 4 November 2026, after the classroom data collection had ended, measured fault-switch responses at accessible ECU test points with connectors fitted and ignition ON. These supplementary observations are reported separately in Table 7(c); they were not used to change the original student assessment scores and cannot retrospectively establish the electrical mechanism of the fault assessed in August.

Instructional Procedures

As detailed in Table 2, the classes received four weekly instructional sessions covering the same objectives (TP 1–TP 5). Experimental students worked in seven groups of five using one trainer, a multimeter, scan tool, the practicum guide, and worksheets. A school technician prepared the trainer before each meeting. Whole-class demonstrations were followed by scheduled practical blocks of 45, 55, 70, and 70 minutes. Divided among seven groups, these blocks implied approximately 6–10 minutes of potential workstation access per group per meeting, or about 35 minutes per group across the intervention. These are scheduled estimates rather than observed individual exposure: student roles, actual operating time, and waiting-group activities were not logged.

Worksheet 1 concerned sensor identification and terminal measurements. Worksheet 2 involved manual TPS variation and interpretation of ECT-related diagnostic information; the trainer did not produce changing coolant temperature. Worksheet 3 introduced DTCs, live data, scanner operation and fault scenarios. The fourth session used a guided complaint-to-verification ECT troubleshooting task. The control class received lectures, slides, printed materials, written DTC/live-data examples, exercises, and fault cases but no trainer, multimeter, or scan tool. The conditions consequently differed in both hardware access and organization of practical activities; those elements cannot be separated statistically.

Instruments and Assessment Procedures

Five instrument categories were used: material validation, media/hardware validation including U1–U10 inspection criteria, teacher/student practicality questionnaires, a cognitive achievement test, and an observational troubleshooting rubric. Their constructs, item numbers, scoring procedures, and measurement evidence are summarized in Table 4. Expert validity and practical ratings were calculated as percentages of maximum possible scores; the detailed expert indicators are preserved in Supplementary Table S1.

Table 4. Research instruments, scoring procedures, and available measurement evidence

Instrument	Construct and indicators	Items	Scoring	Validation and reliability
Material validation sheet	Concept accuracy; curriculum fit; system representation; completeness; fault scenarios	20	Likert 1–5; % of maximum	3 experts; indicator-level results in Supplementary Table S1

Instrument	Construct and indicators	Items	Scoring	Validation and reliability
Media/hardware validation sheet	Hardware engineering; safety and construction; panel and readability; system function; OBD-II diagnosis; operation and maintenance; manual	37 + U1–U10	Likert 1–5; U1–U10 met/not met	3 experts; Tables 6 and 8
Practicality questionnaires	Ease of use; time efficiency; learning benefit; diagnostic skills; technical quality (teacher and student versions)	16 each	Likert 1–5; % of maximum	Content validation not documented; Cronbach's α (students, n = 15) = −0.41
Cognitive test	Five content areas, C1–C5 (blueprint in text)	30	3 points/correct; 0–90	Content: 2 experts (92.61%); Cronbach's α = 0.883, KR-20 = 0.885 (pretest, n = 70); difficulty 0.63–0.99; discrimination D = 0.00–0.74
Troubleshooting rubric	Preparation (8), symptom identification (12), scanner diagnosis (12), component measurement (40), analysis and conclusion (16), work safety (12)	25 sub-skills	4-level descriptors; total 0–100	2 observers; each student scored by both during group work; final score = mean; inter-observer agreement not computed (observer-level scores not retained)

The cognitive test comprised 30 five-option questions on basic EMS concepts (items 1–6; C1–C2), sensors (7–12; C1–C2), actuators and signal flow (13–18; C2–C3), sensor and component diagnosis (19–26; C3–C4), and evaluation of troubleshooting procedures (27–30; C4–C5). Example items used TPS scanner readings under a hesitation complaint and a high ECT reading under a cold-start complaint. These written scenarios were not claimed to replicate the trainer's missing thermal dynamics. Two experts evaluated the item blueprint and wording in May 2026. Their feedback led to revisions of distractors, units, technical terminology, DTC notation, and stimulus length. Without a separate pilot group, item statistics were computed from the pretest responses of all 70 students.

Each correct answer received three recorded points, for a maximum of 90. This is a linear transformation of the number-correct result (percentage = score/0.9) and

was retained to match original school records. The same 30 items were used at pretest and posttest separated by 28 days; test familiarity therefore remained possible. Student practicality questionnaire scores had Cronbach's $\alpha = -0.41$ ($n = 15$), precluding interpretation of the sum as a reliable unidimensional scale. The cognitive test produced $\alpha = 0.883$ and KR-20 = 0.885 but had many easy items, requiring cautious interpretation.

The 25-subskill troubleshooting rubric allocated 8 points to preparation, 12 to symptom identification, 12 to scanner diagnosis, 40 to component measurement, 16 to analysis and conclusion, and 12 to work safety. Each subskill used four performance levels. Table 5 preserves five representative original descriptors for the practised ECT scenario, including the assessment's interpretation of 0–0.12 V as a signal short to ground. That scoring criterion is an original assessment record rather than independently verified evidence of the switch's electrical mechanism.

Table 5. Descriptors of the troubleshooting rubric for five sub-skills (ECT signal-circuit fault scenario)

Sub-skill (max. 4)	Level 4	Level 3	Level 2	Level 1
3.3 Read and interpret DTC	Reads the ECT DTC (P0115 on this trainer) and states its meaning	Reads the code but cannot state its meaning	Misreads the code	Cannot read the code
4.3 Measure signal voltage (connector on, ignition ON)	0–0.12 V measured and interpreted as a low signal caused by a short to ground	Low voltage measured but interpretation incorrect	Measures at the wrong terminal	Does not measure
4.5 Check continuity of sensor-to-ECU line	Finds the abnormal connection of the ECT signal line to ground	Finds the short symptom; location uncertain	Wrong terminal or line	Does not measure
5.1 Determine cause	ECT signal line shorted to ground	Suspects the ECT sensor without circuit evidence	Suspects the ECU without verification	Cannot determine a cause
5.4 Verify after repair	MIL off and sensor data normal	Fault persists	A new fault appears	Does not verify

During the fourth meeting, seven groups completed the ECT case at the trainer, with approximately ten minutes scheduled for each group. Two observers rated each student while group members could watch, discuss and assist one another. Scores were averaged across observers, but separate ratings were not retained and observer agreement could not be quantified. 'Repair' meant restoring the switch to its normal position, clearing the DTC and checking available indications. The recorded individual scores therefore reflect contributions during a collaborative, practised task rather than completion of an independent diagnostic test.

Data Analysis

Validity and practicality were reported as (obtained score / maximum possible score) × 100 and classified following criteria adapted from Sugiyono (2024): 85.01–100% very valid/practical, 70.01–85.00% valid/practical, 50.01–70.00% fairly valid/practical, and ≤50% not valid/practical. Because the student questionnaires accounted for 1,200 of 1,440 possible pooled practicality points (83.33%), teacher and student results were reported separately as well as pooled. Questionnaire results were treated as descriptive ratings.

Cognitive outcomes were summarized using means, SDs and 95% confidence intervals. Shapiro–Wilk and Levene tests examined distributional assumptions. The planned ANCOVA modeled posttest with class and pretest covariates, following Vickers and Altman (2001). The additive-model residuals were approximately normal ($W = 0.966$, $p = 0.056$), but Levene's test ($F = 11.974$, $p < 0.001$) and the class × pretest interaction ($F = 12.681$, $p < 0.001$) indicated unequal variances and nonparallel slopes. Therefore, a common adjusted class difference was not interpreted as a uniform intervention effect.

Wilcoxon signed-rank tests assessed within-class change, and Mann–Whitney U tests compared pretest, posttest, raw gain and normalized gain. Nonparametric effect size was $r = |Z|/\sqrt{N}$ (Fritz et al., 2012). Sensitivity analyses included Hodges–Lehmann estimates, 10,000-resample percentile bootstrap intervals, HC3 pretest-adjusted regressions with and without class × pretest interaction, conditional estimates at selected pretest levels, and comparisons of 26 student pairs with identical pretest scores. HC3 estimation followed Hayes and Cai (2007) and Long and Ervin (2000), while interaction probing followed Hayes and Matthes (2009). Pair-selection details where multiple exact-score matches were possible were unavailable; those estimates were exploratory.

Gain-score and baseline-adjusted analyses estimate different quantities when classes begin at different levels (Lord, 1967). N-Gain was (posttest − pretest)/(90 − pretest), categorized using Hake (1998). Seven experimental and three control students began at 90 and were excluded from N-Gain because the denominator was zero (Marx & Cummings, 2007); all remained in raw-gain analyses with a gain of zero. The significance level was .05. Troubleshooting outcomes were descriptive for the experimental class only. Given one nonrandomized class per condition, none of the student-level analyses removes class–treatment confounding (Kim & Steiner, 2016).

Ethical and Institutional Considerations

The study received administrative permissions but no independent research ethics committee approval. Classroom implementation was conducted under a Faculty of Engineering, Universitas Negeri Padang research permit valid 13 July–13 August 2026, accepted by SMK Negeri 1 Bangkinang; the pretest, posttest, practicality and troubleshooting assessments took place during that period. Teacher and student needs-analysis questionnaires were administered January–February 2026 during routine teaching activities before the faculty permit and without contemporaneous written school authorization. Adult experts voluntarily completed validation forms between April and June 2026, outside the permit's dates.

On 4 November 2026, after the classroom data collection had been completed, the school principal issued retrospective statement No. 070/SMKN1BKN/2026/739. The statement confirmed that the needs analysis and classroom activities had taken

place with the school's knowledge and authorized the use of anonymized data for the thesis and scientific publication. Because it was issued after these activities, the statement constitutes retrospective institutional confirmation and data-use authorization, not prospective research approval, ethics committee approval, or individual informed consent. The participating students were approximately 16–17 years old; written consent from students and parents/guardians was not obtained or documented, and there was no formally granted waiver. These limitations are disclosed rather than represented as resolved by school authorization. Scores were coded before analysis, no student identities are reported, and practical sessions used fuse protection and precautions including ignition OFF before disconnecting connectors.

RESULTS AND DISCUSSION

Needs Analysis and Baseline Conditions

Three teachers identified limited facilities and lecture-dominated instruction (mean ratings 5.00) and rated access to interactive EMS media, school ownership of a trainer, an adequate trainer–student ratio, and visualization of sensor–ECU–actuator flow at 1.00. Among 64 student respondents, self-rated understanding was 1.58 and the attractiveness of current media 1.72; difficulty visualizing sensor and ECU operation was 4.86 and the perceived need for direct practice was 5.00. These findings informed requirements for a reusable panel with visible wiring, accessible terminals, controllable faults and OBD-II diagnosis.

At the July pretest, 17 control students (48.57%) and 11 experimental students (31.43%) were recorded as below the school's mastery threshold of 75. The school threshold and test scores, however, were not documented on a consistent 0–90 or 0–100 scale; the transformation convention is unknown. The counts are therefore retained as contextual school classifications and are not used to establish comparative mastery.

Learning Package Development and Diagnostic Facilities

The researchers refined the practicum guide, worksheet descriptions of terminals and reference values, reset procedures, graded DTC fault scenarios, and shared terminology following expert feedback. The commercial trainer was functionally inspected and its ECU wiring rechecked; no new trainer hardware was fabricated. Figure 3 identifies the panel and its installed components. Figure 4 illustrates test-point sockets, fault switches, the OBD-II connector and the diagnostic display, and Figure 5 maps functional diagnostic paths and conditions used for ECT measurement. The latter is a functional diagram and not a verified schematic of hidden internal fault-switch wiring.

Technical Function and Safety Inspection

All three media/hardware experts marked each of the ten inspection criteria U1–U10 as met, a total of 30 positive judgments out of 30 (Table 6). These are binary judgments covering supply, fuse and ground checks, protection of terminals and ignition components, readable test points, fault indication, diagnostic communication and approximately ten minutes of operation. No numerical measurements of supply stability, fuse trips or durability were recorded in these inspections, so the result should not be interpreted as formal electrical safety certification.

Table 6. Inspection-based function and safety checks (U1–U10) by three media/hardware experts

Code	Test criterion (as recorded in the validation instrument)	Type of evidence	Experts: met
U1	Supply voltage to the ECU/ECM and main components matches the trainer specification	Inspection	3/3
U2	Fuse rating appropriate; circuit is interrupted when the fuse is removed or blown	Inspection	3/3
U3	Ground continuity between panel ground and ECU ground	Inspection	3/3
U4	No supply-to-ground short circuit with ignition OFF	Inspection	3/3
U5	Live terminals and conductors are protected against contact	Inspection	3/3
U6	Ignition coils and their connections are protected; sparks occur only where intended	Inspection	3/3
U7	Signals at main test points (e.g., ECT, TPS) are readable and change with conditions	Inspection	3/3
U8	Fault simulation ON: symptom/DTC appears; restored: symptom/DTC disappears	Inspection	3/3
U9	Scan tool connects to the DLC and reads DTCs and live data	Inspection	3/3
U10	Trainer operates stably for about 10 minutes without overheating or malfunction	Inspection	3/3

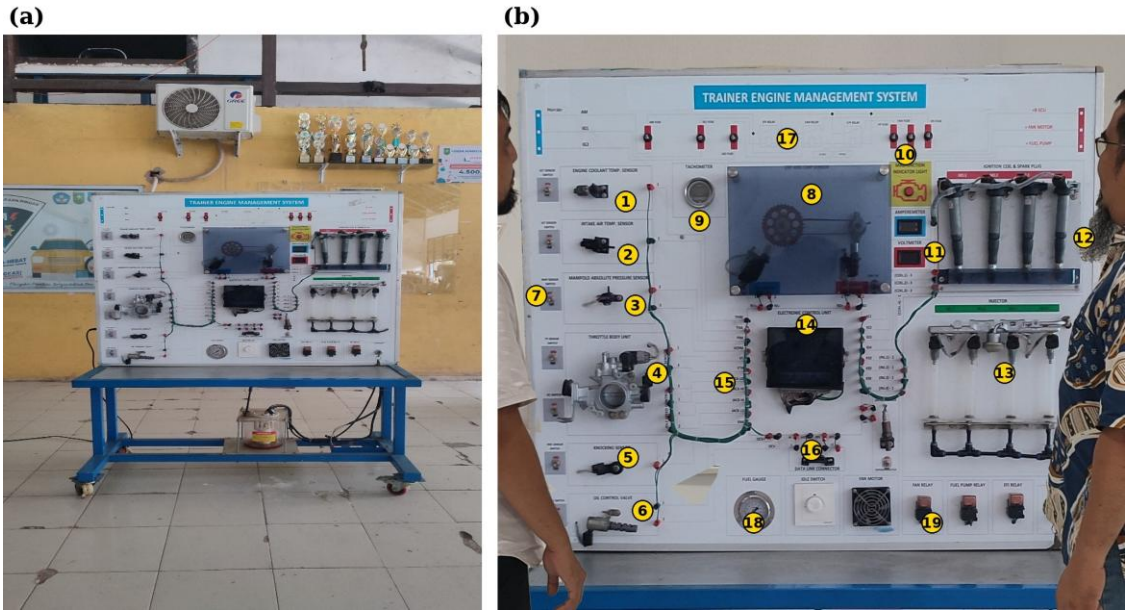


Figure 3. (a) The EMS trainer on its mobile frame; (b) labelled panel: 1 ECT sensor, 2 IAT sensor, 3 MAP sensor, 4 throttle body with TPS and ISC valve, 5 knock sensor, 6 oil control valve, 7 fault-simulation switches (one per circuit, left column), 8 CKP/CMP reluctor assembly, 9 tachometer, 10 MIL, 11 ammeter and voltmeter, 12 ignition coils, 13 injectors with transparent spray tubes, 14 ECU, 15 ECU terminal test points, 16 DLC (OBD-II), 17 fuses and relay circuit, 18 fuel-pressure gauge, 19 fan, fuel-pump, and EFI relays

Table 7 distinguishes three types of evidence. Panel (a) reports behaviour documented through scan-tool screenshots in the trainer manual and supported

qualitatively by U8–U9 expert judgments. It describes DTC indications for ECT (P0115), IAT (P0110), MAP (P0105), TPS (P0120), ISC (P0505) and OCV (P0010), as well as clearing codes, spark, injection, and tachometer function. No knock DTC occurred because no knock was generated. Importantly, the OCV DTC described by the manual was not reproduced in the subsequent technical check.

Panel (b) retains student measurements from the ECT assessment: reference/pull-up voltage of 4.87–5.12 V at THW with the sensor connector disconnected and ignition ON, and 0–0.12 V with the connector connected, ignition ON and ECT fault switch OFF. Because these readings were made under different connector conditions, they should not be interpreted as matched normal-versus-fault observations. Panel (c) presents single DC multimeter readings from the separately conducted post-implementation technical inspection, with connectors fitted and the engine not cranked; these measurements were not part of the student assessment.

Table 7. Functional behaviour and measured signal values, by source of evidence

Circuit or function	Condition	Expected	Observed
(a) Documented in the trainer manual (scan-tool screenshots); confirmed qualitatively by validators (U8–U9)			
ECT	ECT switch OFF; scan tool DTC Analysis	DTC, MIL on	P0115; MIL on
IAT	IAT switch OFF	DTC, MIL on	P0110
MAP	MAP switch OFF	DTC, MIL on	P0105
TPS	TP switch OFF	DTC, MIL on	P0120
ISC valve	ISC switch OFF	DTC, MIL on	P0505
OCV	OCV switch OFF	DTC, MIL on	P0010 (manual); no DTC in the post-study check
Knock	KNK switch OFF	–	No DTC (no knock)
Reset	Switches ON; Erase in scan tool	No DTC, MIL off	As expected
Running mode	Key START; speed regulator turned	Rotation, spark, injection	Spark; injector spray; rpm on tachometer
(b) Measured by students during the study (ECT fault scenario; test point THW to panel ground, ignition ON)			
ECT reference voltage	Sensor connector disconnected	5 V ± 0.2 V	4.87–5.12 V
ECT signal voltage	Connector connected; ECT switch OFF	Low (≈ 0 V) if shorted to ground	0–0.12 V
ECT resistance, injector supply, normal-condition signals	–	Rubric reference values	Not recorded
(c) Post-implementation technical measurements (4 November 2026; ECU test point to panel ground, ignition ON, engine not cranked, all connectors fitted)			
Circuit (test point)	*Switch ON (V)*	*Switch OFF (V)*	*DTC displayed*
ECT (THW)	2.380	2.391	P0115

Circuit or function	Condition	Expected	Observed
IAT (THA)	1.948	1.964	P0110
MAP signal (PIM)	3.644	3.63	P0105
MAP supply (VCPM)	5.04	5.03	P0105
TPS signal (VTH)	0.618	Not recorded	P0120
TPS supply (VC)	5.05	0.687	None
ISC (IACA/IACB)	0.617	4.65	P0505
OCV (OCV+)	1.114	1.051	None

In the post-implementation technical measurements (Table 7c), THW, THA, PIM, VCPM and OCV+ changed by less than 0.07 V when fault switches were moved from ON to OFF, although the ECT, IAT and MAP switches still produced the recorded DTCs. Larger changes were observed at the TPS supply VC (5.050 to 0.687 V) and ISC drive IACA/IACB (0.617 to 4.650 V). These observations demonstrate differing terminal responses but do not establish internal short/open mechanisms. In particular, OCV produced no DTC in this check despite the manual documenting P0010.

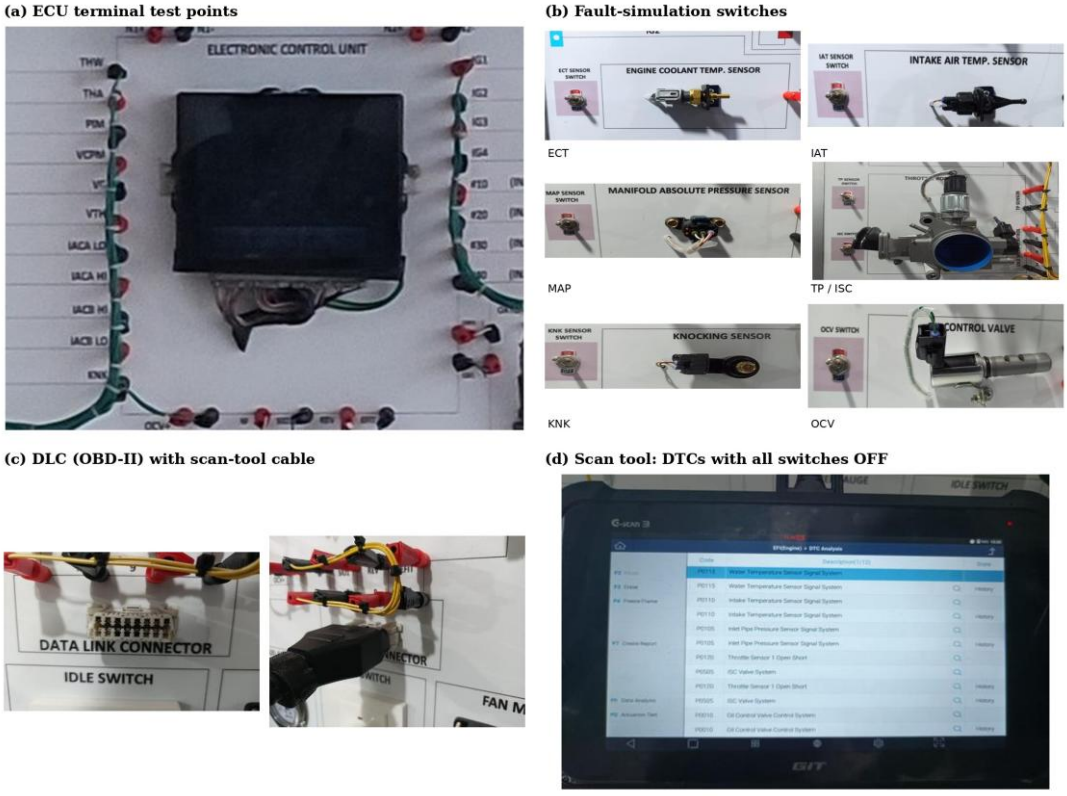


Figure 4. Diagnostic facilities of the trainer: (a) ECU terminal test points (banana sockets labelled with ECU terminal names); (b) fault-simulation switches (ON/OFF) beside the ECT, IAT, MAP, TP/ISC, knock, and OCV circuits; (c) data link connector without and with the scan-tool cable; (d) scan-tool DTC list with all fault switches OFF (photographs from the trainer manual and the study)

The ECT result differed between measurement sessions. Students recorded 0–0.12 V at THW on 6 August 2026; the subsequent switch-OFF check recorded 2.391 V. The discrepancy cannot be resolved from the retained record and may involve changes in setup, equipment or recording, none of which was verified. Therefore, the

August reading is retained as an original assessment observation and rubric value, not definitive proof of an ECT signal short to ground. Overall, DTC generation was better documented than fault-dependent test-point changes for several circuits.

Expert Validity and User-Perceived Practicality

Expert assessments were 91.33% for material content (274/300), 92.79% for media/hardware (515/555) and 92.61% for the learning/test instrument (213/230), all in the very valid category. The more detailed expert ratings are preserved in Supplementary Table S1. Relatively lower-scoring indicators concerned the ECU's role, alignment of troubleshooting with the curriculum, operation and maintenance, hardware engineering, and time allocation for testing. One material item referred to an oxygen sensor and vehicle speed sensor, which were not fitted to the trainer. Removing that item changed the material percentage to 91.58% (261/285), without changing its classification. Validators recommended clearer cable identification, reference-value and test-point guidance, pre/post-use checklists, more graded fault scenarios, and clearer distractors.

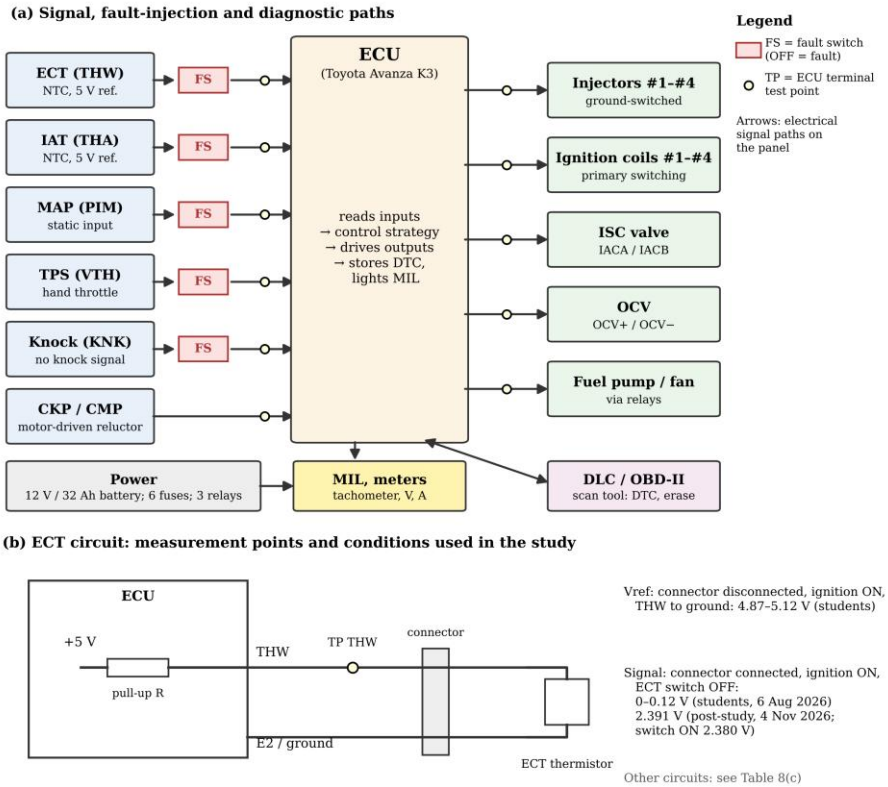


Figure 5. (a) Signal, fault-injection, and diagnostic paths of the trainer; (b) ECT circuit with the measurement point and the conditions under which the reference and signal voltages were measured

Teachers rated practicality at 93.33% (224/240) and 15 students at 90.50% (1,086/1,200), yielding a pooled 90.97% (1,310/1,440; Table 8). All three teachers and 14 students were in the very practical category; one student was in the practical category. Teacher aspect ratings ranged from 86.67% to 100%, and student aspect ratings from 86.00% for measurement/diagnosis to 92.67% at the high end. Because students accounted for 83.33% of the pooled maximum score, the pooled percentage is not an equally weighted teacher–student assessment.

The student practicality questionnaire had negative internal consistency (Cronbach's α = −0.41, n = 15). Favourable ratings are therefore reported descriptively rather than as a valid internally consistent total scale; coding, item covariance and dimensionality cannot be resolved from the aggregate responses. Content validation of practicality questionnaires was not documented.

Table 8. Expert-judged validity and user-perceived practicality of the EMS diagnostic learning package

Evaluation dimension	Respondents	Score / maximum	Percentage	Category
Panel A — Expert validity				
Material	3 experts	274 / 300	91.33%	Very valid
Media/hardware	3 experts	515 / 555	92.79%	Very valid
Learning/test instrument	2 experts	213 / 230	92.61%	Very valid
Panel B — Perceived practicality				
Teachers	3	224 / 240	93.33%	Very practical
Students (XI TKR 2)	15	1,086 / 1,200	90.50%	Very practical
Pooled	18	1,310 / 1,440	90.97%	Very practical

The cognitive test showed high internal consistency (α = 0.883; KR-20 = 0.885) and significant item–total correlations for 23 of 30 items. However, 25 items were easy (difficulty > .70), five medium and none difficult; discrimination was very good for two, good for 15, fair for nine, poor for three (items 3, 6 and 9), and unacceptable for item 1 (D = 0.00). Eighteen items assessed C1–C3. Thus, the test was more sensitive to basic conceptual understanding than to advanced diagnostic reasoning, and high reliability alone does not prove construct validity (Taber, 2018).

Cognitive Learning Outcomes

Both classes improved from pretest to posttest (Wilcoxon $p < .001$ within each class; Table 9, Panel A). Experimental-class means increased from 77.74 (SD 9.72) to 86.06 (SD 4.29), a gain of 8.31; control-class means rose from 67.03 (SD 20.31) to 79.80 (SD 11.02), a gain of 12.77. Initial class means differed by 10.71 points (Mann–Whitney $p = .042$) and the control class had substantially greater variance. Seven experimental and three control students scored at the test maximum of 90 at pretest; by posttest, 13 experimental and ten control students were at 90. Figure 6a displays individual trajectories and class means.

Between-class mean differences were 6.26 posttest points, −4.46 raw-gain points and 0.069 normalized-gain units (experimental minus control; Table 9, Panel B). Mann–Whitney comparisons of gain ($p = .102$; $r = .196$) and N-Gain ($p = .074$; $r = .230$) were not statistically significant. The bootstrap CI for the mean gain difference [−8.31, −0.77] did exclude zero, whereas the Hodges–Lehmann estimate was −3.0 [−6.0, 0.0] and the rank test was not significant. These procedures address different distributional quantities; the Mann–Whitney test is not simply a median test when spreads differ (Hart, 2001). Neither unadjusted result eliminates baseline imbalance.

A common-slope HC3 adjusted posttest comparison estimated 0.82 points [−0.25, 1.90], $p = .131$. However, the class × pretest interaction was significant ($p = .001$), and the fitted slopes were 0.53 in control and 0.39 in experimental (Figure 6b); thus, the additive coefficient cannot be interpreted as a uniform treatment effect. Exploratory conditional class differences were 2.05 [0.60, 3.50] at pretest 67 and −0.48 [−1.59, 0.64] at pretest 85. Among 26 score-matched pairs, gains were 9.00 and 7.96, p

= .110. These conditional and matched comparisons are descriptive sensitivity analyses, not independent estimates of treatment causality.

Table 9. Cognitive learning outcomes: Descriptive statistics, between-class comparisons, and sensitivity analyses

Class / comparison	Pretest / test statistic	Posttest / p	Gain / r	N-Gain / difference [95% CI]
Panel A — Descriptive statistics				
Experimental (n = 35)	77.74 (9.72) [74.40, 81.08]	86.06 (4.29) [84.58, 87.53]	8.31 [6.18, 10.45]	0.720 (0.195) [0.645, 0.796]; n = 28
Control (n = 35)	67.03 (20.31) [60.05, 74.01]	79.80 (11.02) [76.02, 83.58]	12.77 [9.44, 16.10]	0.651 (0.209) [0.576, 0.727]; n = 32
Panel B — Between-class comparisons (experimental − control)				
Pretest	U = 439.5	0.042	r = 0.243	10.71 [3.60, 18.34]†
Posttest	U = 423.5	0.023	r = 0.271	6.26 [2.49, 10.20]†
Raw gain	U = 751.0	0.102	r = 0.196	−4.46 [−8.31, −0.77]†; HL −3.0 [−6.0, 0.0]
N-Gain (n = 28 vs 32)	U = 328.0	0.074	r = 0.230	0.069 [−0.032, 0.171]†
Adjusted posttest (HC3 additive)	t(67) = 1.53	0.131	—	0.82 [−0.25, 1.90]
Class × pretest interaction	t(66) = −3.39	0.001	—	Slope difference −0.14
Conditional at pretest 67	—	0.006	—	2.05 [0.60, 3.50]
Conditional at pretest 85	—	0.395	—	−0.48 [−1.59, 0.64]
Matched 26 identical-score pairs	Wilcoxon	0.110	—	Gain 9.00 vs 7.96

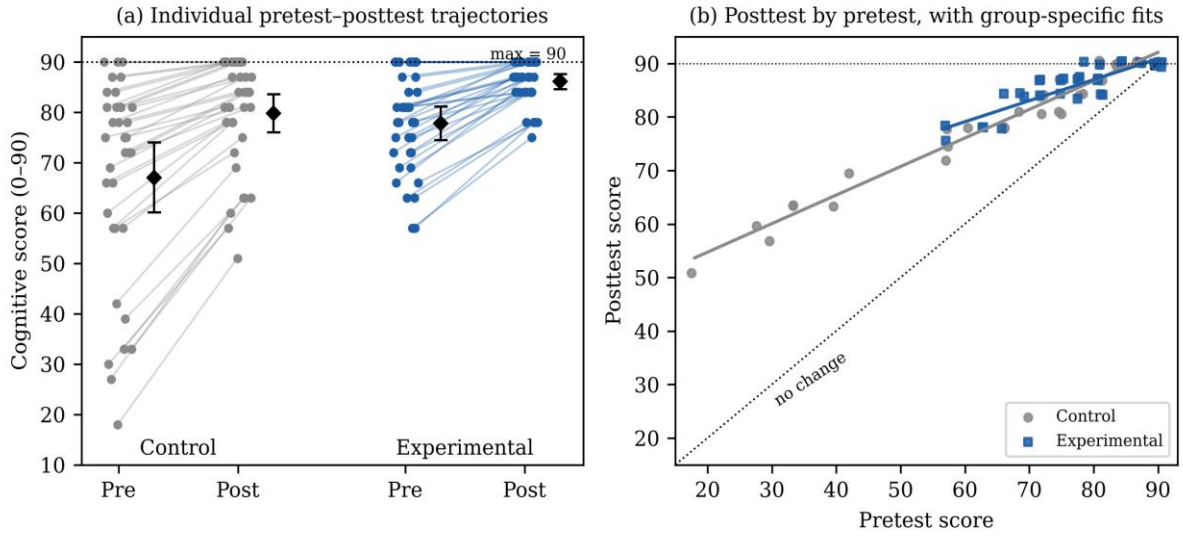


Figure 6. (a) Individual pretest–posttest scores with class means and 95% CIs; (b) posttest by pretest with group-specific least-squares lines (points jittered slightly for visibility)

Both classes improved, but the rank-based gain comparisons did not establish a comparative advantage for trainer-supported instruction. Baseline differences, the upper score limit, easy and repeated test items, and the absence of replicated classes per condition restrict inference; neither a causal difference nor equivalence can be concluded.

Observed Troubleshooting Performance

The experimental class averaged 88.60 (SD 5.59; range 78.50–98.50) on the observed ECT troubleshooting rubric. Twenty-four students (68.57%) were categorized very skilled and eleven (31.43%) skilled. Table 10 presents aspect means, ranging from component measurement (87.82%) to work safety (91.07%). All seven groups completed the rubric-defined cause and restoration/verification steps, but the fault mechanism remained uncertain because the low ECT reading in August was not reproduced during the subsequent technical check.

Table 10. Observed troubleshooting performance by aspect (experimental class, n = 35)

Aspect	Maximum score	Mean score	Percentage (%)
Preparation	8	7.20	90.00
Symptom identification	12	10.64	88.69
Scanner diagnosis	12	10.57	88.10
Component measurement	40	35.13	87.82
Analysis and conclusion	16	14.13	88.30
Work safety	12	10.93	91.07
Total	100	88.60 (SD 5.59)	–

The high recorded scores document participation in a structured group task rather than independent diagnostic competence or transfer. Students could assist peers, and individual exposure to the single trainer was not measured. The lack of observer-level data also prevents precision claims about small individual score differences.

DISCUSSION

Technical Performance and Task-Specific Fidelity

The trainer enables physical interaction with selected electrical signals, ECU-driven actuators, diagnostic trouble codes (DTCs), the MIL, and OBD-II communication. This offers task-relevant functional fidelity, although fidelity should be judged by the cues and actions needed for instruction rather than physical resemblance to a complete engine (Hamstra et al., 2014). Physical and virtual laboratory research likewise cautions that learning value depends on instructional tasks and the evidence students collect and interpret (de Jong et al., 2013; Zacharia & Olympiou, 2011). The motor-driven CKP/CMP assembly produced visible sparks and injector spray, while combustion, vacuum, dynamic coolant and intake temperatures, knock, oxygen sensing and closed-loop control remained absent. The trainer therefore reproduced selected electrical diagnostic functions, not full engine operating conditions.

An especially important finding was the distinction between DTC feedback and measurable electrical evidence. The trainer manual documented DTC responses for six switches, but the post-implementation measurements showed that activating ECT, IAT and MAP faults produced their documented codes without substantial voltage changes at the accessible THW, THA and PIM test points (Table 7). The TPS supply

and ISC drive did exhibit sizable changes, but single DC readings did not establish the exact fault mechanism or ECU waveform. The discrepancy between the August ECT assessment voltage (0–0.12 V) and the later measured value (2.391 V) remained unresolved; P0010 for OCV was documented in the manual but not reproduced post-study. The findings therefore do not establish that an activated switch consistently creates an electrically localizable short or open at the measured point.

Engineering investigations support separating injected fault conditions from their observable effects. Lee et al. (2022) analysed simulated EFI fault-mode–effect relationships, and Ghimire et al. (2018) derived relationships between injected disturbances and sensor measurements in electric power steering. These studies do not validate this trainer's wiring; rather, they support a task-specific evaluation of what a diagnostic device actually makes observable. DTC retrieval can be practised wherever the code is reproduced, but terminal-voltage fault localization should be taught only where reproducible responses and wiring information support it. Expert pass/fail inspection and the absence of longitudinal use logs do not establish equipment durability or formal electrical safety.

Instructional Design and Contribution to Diagnostic Learning

The present contribution differs from a claim of hardware invention. Digital EMS media described by Fitriyanto et al. (2020) and Mubasir (2020) allow representations of circuit and scan-tool information; physical training devices studied by Handoyono and Rabiman (2023) and Tohadi et al. (2020) make particular components available in an educational setting. More recent Hidayat et al. (2024) and Nanda et al. (2025) studies illustrate other EFI simulator and mobile teaching-factory approaches, while Liao et al. (2023) examined virtual-reality-supported practical vehicle-electrical instruction. Their technical functions, accompanying pedagogies and outcome measures differ, so they do not establish that one medium is inherently more effective.

The distinguishing design element here is the alignment of the commercial trainer's diagnostic affordances with a written instructional sequence: component identification, reference and signal checks, DTC retrieval, hypothesis formation, evidence evaluation, and restoration/verification. That sequence operationalizes structured troubleshooting principles (Jonassen, 2000; Jonassen & Hung, 2006; Schaafstal et al., 2000) and offers opportunities for experience, reflection and hypothesis testing within experiential learning (Kolb & Kolb, 2005; Morris, 2020). Liang (2020) also illustrates the value of organized automotive diagnostic processes and knowledge in collaborative professional settings. However, the study did not directly measure the hypothesized cognitive mechanisms; following structured steps is not evidence that transferable reasoning developed. The technical observations help specify instructional boundaries. For fault switches that consistently generate DTCs but show little measured change at the accessible terminal, the package can support diagnostic-code interpretation and discussion of alternative fault hypotheses, but cannot claim voltage-confirmed fault localization. Even where voltage changes were observed, the mechanism requires further verification. The package is therefore a documented application of established diagnostic-learning principles to a commercial EMS device, not a novel general theory of diagnosis.

Expert Validity, Practicality, and Measurement Quality

Experts rated the material, media/hardware, and learning/test instruments above 91%, and all U1–U10 inspection criteria were marked met. These results support

local alignment with content and technical inspection requirements, consistent with educational-media selection considerations (Arsyad, 2013). They are not direct evidence of comparative learning effects, formal safety certification or durable routine operation (Clark, 1983). The relatively weaker expert items concerned ECU signal processing, troubleshooting alignment, maintenance guidance and test duration. The original material score also included one item about absent oxygen and speed sensors; excluding it left the classification unchanged but underscores the importance of item-level scrutiny.

The teacher (93.33%) and student (90.50%) practicality ratings were favourable, yet the student instrument had Cronbach's $\alpha = -0.41$ with only 15 respondents. Thus, the pooled 90.97% describes favourable responses, not a validated practicality scale. The comparatively lower student measurement/diagnosis rating, teacher time/preparation ratings, and lower component-measurement rubric percentage (87.82%) identify a recurring area in which clearer reference values, more individual access and better instructions may help; they do not prove one common cause. Novelty or acquiescent responding cannot be ruled out. This separation between rated product suitability and learning outcomes is essential to avoid attributing cognitive benefits to medium ratings alone (Clark, 1983).

Cognitive Outcomes and Comparative Effectiveness

Both classes improved significantly, but they began at different levels. The control class started 10.71 points below the experimental class and with more dispersed results, leaving greater possible room for raw gain and vulnerability to regression toward the mean (Barnett et al., 2005). Its greater mean raw gain (12.77 versus 8.31) cannot therefore establish conventional instruction's superiority. Conversely, the experimental class's higher posttest mean was smaller than its pretest advantage and cannot establish a trainer advantage.

A pronounced score ceiling further limited discrimination: seven experimental and three control students already scored 90 at pretest, increasing to 13 and ten at posttest. Restricted outcome range can distort comparisons (Austin & Brunner, 2003). Unequal exclusions from N-Gain removed more experimental students whose normalized gains were mathematically undefined, so the direction of any resulting bias is uncertain. Although the class × pretest interaction was significant, the available analysis does not prove that ceiling effects caused the slope difference; conditional effects at chosen pretest values are exploratory descriptions.

The cognitive measure also did not closely match the intervention's higher-order aims. Twenty-five of 30 items were easy, eighteen targeted C1–C3, and four had poor or zero discrimination. High internal consistency does not establish diagnostic reasoning validity. Conventional instruction covered the same concepts and written DTC cases, so both classes could learn what the test emphasized. Reusing the same items after 28 days could also contribute to practice or sensitization effects (Willson & Putnam, 1982). These are plausible measurement explanations rather than demonstrated causes of the nonsignificant comparisons.

The Mann–Whitney comparisons of raw gain ($p = .102$) and N-Gain ($p = .074$) did not show a detectable difference. The bootstrap interval for the mean raw gain difference excluded zero, whereas the rank comparison and Hodges–Lehmann interval did not; these methods evaluate distinct quantities and are not directly interchangeable. The common-slope HC3 adjusted difference of 0.82 [−0.25, 1.90] was

also inconclusive and, because of the class × pretest interaction, was not a uniform effect. The sensitivity analyses and score-matched comparisons cannot remove the fundamental one-class-per-condition confounding.

Other simulator or trainer studies have reported favourable learning effects (Perdana Sutrisno et al., 2025; Rohayati et al., 2025; Taufiq et al., 2024; Tohadi et al., 2020), as have studies coupling EFI and vehicle-electrical simulation with explicit instructional strategies (Liao et al., 2023; Nanda et al., 2025). Differences across technical systems, teaching strategies, outcome constructs and allocation designs preclude treating their results as direct replications of this setting. More generally, the learning consequences of a medium depend on teaching activities and what assessments capture (Brinson, 2015; Clark, 1983). With two intact classes taught at different times by the researcher, the observed data neither establish causal superiority nor demonstrate equivalence.

Observed Troubleshooting and Transfer

The experimental class averaged 88.60 on the group-based ECT assessment, with most scores in the very skilled category. Students completed the prescribed symptom-to-DTC-to-measurement-to-verification procedure on the practised scenario. This resembles structured troubleshooting processes that can be taught through systematic cognitive task analysis (Schaafstal et al., 2000) and includes opportunities for reflection described by Morris (2020). Nevertheless, the later ECT electrical discrepancy weakens any claim that students verified an actual short to ground rather than matching the rubric's specified interpretation.

Each group had approximately ten minutes at the station during assessment and could exchange information. Students had already practised related faults. The resulting individual ratings therefore cannot distinguish procedural familiarity, observation of peers, and independent diagnostic reasoning. No troubleshooting baseline, no control-class practical task, no unfamiliar fault and no retained observer-level ratings were available; agreement and individual precision could not be assessed. Evidence about transfer requires performance on new tasks rather than repeated instructional cases (Blume et al., 2010). Research on car-mechatronics diagnosis underscores the value of measurement choice and sequence when assessing problem-solving behaviour (Abele, 2018; Hartmann et al., 2023).

Scientific Contributions, Implications, and Limitations

Three contributions emerge. First, the study documents how a commercial EMS trainer can be embedded in a structured instructional package with worksheets, measurement references, DTC cases and restoration checks. Second, it demonstrates why technical validation of diagnostic learning equipment must distinguish scan-tool feedback from reproducible fault-related electrical measurements. Third, it provides an evaluation that keeps expert validity, perceived practicality, cognitive outcomes, and observed group-based troubleshooting separate, rather than equating product approval with effectiveness. These are contributions to the specification and responsible evaluation of diagnostic learning environments, not proof of a new causal pedagogy.

Teachers can apply the sequence using circuits whose relevant diagnostic response is documented, while reserving measurement-based fault localization for circuits with independently verified, repeatable voltage and wiring responses. Schools should improve reference tables, pre-use checklists, scenario diversity, equipment

logs, and equitable individual operating opportunities. Future evaluations should include replicated or randomized classes, sensitive higher-order tests and independent unfamiliar-fault assessments.

Limitations include a single school; one nonrandomized, nonequivalent class per condition; differing timetables and researcher-as-teacher; ceiling effects and repeated easy cognitive items; small practicality sample with negative internal consistency; one practised group-based troubleshooting case; absent baseline and control troubleshooting data; no observer-agreement records; inspection-only safety checks; incomplete quantitative operating measurements; the unresolved ECT readings and undocumented internal fault wiring; and lack of independent ethics review or documented guardian/student consent. Taken together, the results establish local feasibility of structured DTC-based practice while leaving causal comparative effectiveness and transfer of independent diagnostic competence unestablished.

CONCLUSIONS

This study developed and evaluated a structured diagnostic learning package for a commercially available Engine Management System (EMS) trainer, integrating a practicum guide, worksheets, fault scenarios, and assessment instruments with selected electrical and OBD-II diagnostic functions. Its scientific contribution is the translation of structured troubleshooting into instructional tasks, the documentation of diagnostic feedback versus observable electrical evidence, and the separation of product suitability from measured learning outcomes.

Expert validation was favourable for material (91.33%), media/hardware (92.79%) and learning/test instruments (92.61%); all ten inspection-based function and safety checks were met, although those judgments were not formal electrical-safety certification. Teachers and students gave favourable practicality ratings, interpreted descriptively because the student questionnaire had negative internal consistency. Both classes improved cognitively, but between-class raw-gain ($p = .102$) and N-Gain ($p = .074$) comparisons did not establish an advantage, and baseline imbalance and class–treatment confounding restricted causal inference. The experimental class achieved a mean 88.60 on group-based assessment of one practised ECT fault, which did not establish independent competence or transfer.

Technical observations showed DTC generation without corresponding substantial test-point voltage changes in several circuits, and the ECT assessment voltage could not be reproduced post-study. The package is therefore locally feasible for structured DTC-based diagnostic practice; measurement-confirmed fault localization requires circuit-specific verification, and broader educational impact remains to be established.

RECOMMENDATIONS

Teachers should use a structured progression from component identification and reference measurements to DTC interpretation, fault hypotheses, cause analysis, and verification. They should vary cases beyond rehearsed scenarios and limit voltage-based fault-localization instruction to circuits whose wiring and responses have been independently confirmed. The observed TPS supply and ISC drive changes are candidates for further verification rather than proof of a particular switch mechanism. Schools should provide reference tables, inspection checklists, terminal and DTC

guides, equipment-use logs, and individual operating opportunities when groups share a single trainer.

Future studies should replicate conditions across several schools or classes with randomized or appropriately matched allocation; improve cognitive testing with difficult C4–C6 items, parallel forms and adequate ceiling; assess individual troubleshooting before and after instruction in both groups using unfamiliar faults; record diagnostic accuracy, time and sequences of measurements; estimate inter-observer agreement; and repeatedly measure circuit responses and technical reliability. Research involving minors should obtain prospective ethical review and documented participant and parent/guardian consent or an authorized waiver where legally applicable.

Acknowledgment
The authors thank SMK Negeri 1 Bangkinang, the expert validators, and the teachers and students who participated in this study. During manuscript preparation, the authors used Claude (Anthropic) and ChatGPT (OpenAI) to to enhance the clarity of the writing. The authors retain full responsibility for the accuracy of the scientific content and final manuscript.

Funding Information
This research was partially funded by Universitas Muhammadiyah Purwokerto.

Conflict of Interest Statement
The authors declare no conflict of interest.

Informed Consent
The questionnaires, tests and troubleshooting activities formed part of routine educational activity; nevertheless, written informed consent from the minor students and their parents or legal guardians was not obtained or documented. No waiver was granted. Adult expert validators voluntarily completed online forms. Data were anonymized using participant codes before analysis. The later school statement permitted anonymized academic and publication use but did not constitute prior individual consent.

Ethical Approval
No research ethics committee reviewed or approved this study. Classroom implementation operated under an administrative research permit issued by the Faculty of Engineering, Universitas Negeri Padang (13 July–13 August 2026) and accepted by the school. A later principal's statement (No. 070/SMKN1BKN/2026/739, 4 November 2026) retrospectively confirmed school knowledge of the earlier needs-analysis and classroom activities and permitted de-identified publication. Neither that letter nor the faculty permit constituted prospective ethics approval. The January–February needs analysis and April–June expert validation preceded the classroom permit.

Data Availability
The anonymized pretest–posttest scores, validation and practicality summaries, and troubleshooting scores are available from the corresponding author, R. I. Saputra, upon reasonable request.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Robi Irza Saputra	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓		✓	
Rifdarmon	✓	✓								✓		✓		
Arwizet Karudin		✓								✓				
Toto Sugiarto										✓	✓			

C : **C**onceptualization I : **I**nvestigation Vi : **Vi**sualization
M : **M**ethodology R : **R**esources Su : **Su**pervision

So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
P : **P**roject administration
Fu : **Fu**nding acquisition

REFERENCES

Abele, S. (2018). Diagnostic problem-solving process in professional contexts: Theory and empirical investigation in the context of car mechatronics using computer-generated log-files. *Vocations and Learning, 11*(1), 133–159. https://doi.org/10.1007/s12186-017-9183-x

Alfana, Y. D., Yuliady, I., Zulfadli, Z., & Fernandes, E. (2024). Learning media development of lighting simulator in electrical maintenance training course for vocational education. *Jurnal Pendidikan Teknologi Kejuruan, 7*(4), 276–285. https://doi.org/10.24036/jptk.v7i4.40023

Arsyad, A. (2013). *Media pembelajaran* [Learning media]. Rajawali Pers. https://lib.umpr.ac.id/opac/detail/931

Austin, P. C., & Brunner, L. J. (2003). Type I error inflation in the presence of a ceiling effect. *The American Statistician, 57*(2), 97–104. https://doi.org/10.1198/0003130031450

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology, 34*(1), 215–220. https://doi.org/10.1093/ije/dyh299

Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management, 36*(4), 1065–1105. https://doi.org/10.1177/0149206308321192

Brinson, J. R. (2015). Learning outcome achievement in non-traditional (virtual and remote) versus traditional (hands-on) laboratories: A review of the empirical research. *Computers & Education, 87*, 218–237. https://doi.org/10.1016/j.compedu.2015.07.003

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research, 90*(4), 499–541. https://doi.org/10.3102/0034654320933544

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445–459. https://doi.org/10.3102/00346543053004445

de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science, 340*(6130), 305–308. https://doi.org/10.1126/science.1230579

Feisel, L. D., & Rosa, A. J. (2005). The role of the laboratory in undergraduate engineering education. *Journal of Engineering Education, 94*(1), 121–130. https://doi.org/10.1002/j.2168-9830.2005.tb00833.x

Firmansyah, A. I., & Arizal, H. (2026). Pengembangan media pembelajaran trainer kendaraan listrik di SMK Rajasa Surabaya [Development of an electric vehicle trainer as learning media at SMK Rajasa Surabaya]. *Jurnal Pendidikan Teknik Mesin, 15*(1). https://ejournal.unesa.ac.id/index.php/jurnal-pendidikan-teknik-mesin/article/view/76434

Fitriyanto, P., Wijaya, M. B. R., Septiyanto, A., & Roziqin, A. (2020). Pengembangan multimedia mata pelajaran engine management system untuk meningkatkan kompetensi mengukur besaran listrik actuator pada engine bensin

[Development of engine management system multimedia to improve competence in measuring actuator electrical quantities on petrol engines]. *Jurnal Pendidikan Teknik Mesin, 20*(1). https://doi.org/10.15294/jptm.v20i1.27928

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*(1), 2–18. https://doi.org/10.1037/a0024338

Ghimire, R., Zhang, C., & Pattipati, K. R. (2018). A rough set-theory-based fault-diagnosis method for an electric power-steering system. *IEEE/ASME Transactions on Mechatronics, 23*(5), 2042–2053. https://doi.org/10.1109/TMECH.2018.2863119

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. https://doi.org/10.1007/BF02504676

Gupta, S. L., Mittal, A., Singh, S., & Dash, D. N. (2024). Demand-driven approach of vocational education and training (VET) and experiential learning: A thematic analysis through systematic literature review (SLR). *Asian Education and Development Studies, 13*(1), 45–63. https://doi.org/10.1108/AEDS-07-2023-0083

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74. https://doi.org/10.1119/1.18809

Hamstra, S. J., Brydges, R., Hatala, R., Zendejas, B., & Cook, D. A. (2014). Reconsidering fidelity in simulation-based training. *Academic Medicine, 89*(3), 387–392. https://doi.org/10.1097/ACM.0000000000000130

Handoyono, N. A., & Rabiman, R. (2023). Pengembangan training kit engine management system Toyota Avanza K3-VE untuk praktikum otomotif [Development of a Toyota Avanza K3-VE engine management system training kit for automotive practicum]. *Jurnal Pendidikan Vokasi Otomotif, 6*(1), 37–48. https://doi.org/10.21831/jpvo.v6i1.65886

Hart, A. (2001). Mann-Whitney test is not just a test of medians: Differences in spread can be important. *BMJ, 323*(7309), 391–393. https://doi.org/10.1136/bmj.323.7309.391

Hartmann, S., Güzel, E., & Gschwendtner, T. (2023). Digital measurement of hands-on performance? Ecological validation of a computer-based assessment of automotive repair skills. *Empirical Research in Vocational Education and Training, 15*, Article 15. https://doi.org/10.1186/s40461-023-00153-x

Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods, 39*(4), 709–722. https://doi.org/10.3758/BF03192961

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods, 41*(3), 924–936. https://doi.org/10.3758/BRM.41.3.924

Hidayat, N., Wakhinuddin, Lapisa, R., Milana, Muslim, Sardi, J., & Wirdianto, E. (2024). Development of smart simulator for electronic fuel injection (EFI) fuel system based on quick response code (QR code) for learning media. *International Journal on Informatics Visualization, 8*(3), 1319–1324. https://doi.org/10.62527/joiv.8.3.2209

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development, 48*(4), 63–85. https://doi.org/10.1007/BF02300500

Jonassen, D. H., & Hung, W. (2006). Learning to troubleshoot: A new theory-based design architecture. *Educational Psychology Review, 18*(1), 77–114. https://doi.org/10.1007/s10648-006-9001-8

Kim, Y., & Steiner, P. (2016). Quasi-experimental designs for causal inference. *Educational Psychologist, 51*(3–4), 395–405. https://doi.org/10.1080/00461520.2016.1207177

Kolb, A. Y., & Kolb, D. A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning & Education, 4*(2), 193–212. https://doi.org/10.5465/amle.2005.17268566

Lee, D., Lee, D., & Na, J. (2022). Automatic failure modes and effects analysis of an electronic fuel injection model. *Applied Sciences, 12*(12), Article 6144. https://doi.org/10.3390/app12126144

Liang, J. S. (2020). A process-based automotive troubleshooting service and knowledge management system in collaborative environment. *Robotics and Computer-Integrated Manufacturing, 61*, Article 101836. https://doi.org/10.1016/j.rcim.2019.101836

Liao, C.-W., Liao, H.-K., Chen, B.-S., Tseng, Y.-J., Liao, Y.-H., Wang, I.-C., Ho, W.-S., & Ko, Y.-Y. (2023). Inquiry practice capability and students' learning effectiveness evaluation in strategies of integrating virtual reality into vehicle body electrical system comprehensive maintenance and repair services practice: A case study. *Electronics, 12*(12), Article 2576. https://doi.org/10.3390/electronics12122576

Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician, 54*(3), 217–224. https://doi.org/10.2307/2685594

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 68*(5), 304–305. https://doi.org/10.1037/h0025105

Ma, J., & Nickerson, J. V. (2006). Hands-on, simulated, and remote laboratories: A comparative literature review. *ACM Computing Surveys, 38*(3), Article 7. https://doi.org/10.1145/1132960.1132961

Maksum, H., Purwanto, W., Siman, Ampera, D., Yuvenda, D., & Hasan, H. (2024). Improving problem-solving and communication skills in automotive vocational education through the development of teaching factory model with problem-based learning (TEFA-PBL) concept. *International Journal of Education in Mathematics, Science and Technology, 12*(2), 364–386. https://doi.org/10.46328/ijemst.3941

Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics, 75*(1), 87–91. https://doi.org/10.1119/1.2372468

Morris, T. H. (2020). Experiential learning – a systematic review and revision of Kolb's model. *Interactive Learning Environments, 28*(8), 1064–1077. https://doi.org/10.1080/10494820.2019.1570279

Mubasir, M. M. (2020). Pengembangan multimedia interaktif penggunaan scan tool EFI berbasis Flash pada kompetensi dasar memahami dan memelihara engine management system [Development of Flash-based interactive multimedia on EFI scan-tool use for understanding and maintaining the engine management

system]. *Sainteknol: Jurnal Sains dan Teknologi, 18*(2), 112–120. https://doi.org/10.15294/sainteknol.v18i2.17070

Nanda, I., Simatupang, W., Maksum, H., Rifdarmon, Syaifullah, L., & Putra, R. (2025). Development of innovative mobile QR-EFI simulator in problem-based teaching factory (PBTF) model to enhance students' 4C skills. *International Journal of Interactive Mobile Technologies, 19*(18), 18–31. https://doi.org/10.3991/ijim.v19i18.57285

Perdana Sutrisno, V. L., Pardjono, Wagiran, Nurtanto, M., Ratnawati, D., Marganingsih, A., & Mutohhari, F. (2025). Simulation-based diagnostic learning with diagnostic trouble box (DTB): Enhancing analytical thinking skills in vocational automotive education. *Salud, Ciencia y Tecnología, 5*, Article 1956. https://doi.org/10.56294/saludcyt20251956

Purnama, Y. A., & Wagiran. (2020). The relevance of autotronic engineering competence at vocational high school to the world of work. *Jurnal Pendidikan Vokasi, 10*(3), 300–314. https://doi.org/10.21831/jpv.v10i3.33332

Rohayati, R., Qudus, N., & Widjanarko, D. (2025). The effectiveness of simulator learning media in learning motorcycle charging systems at vocational high schools. *Journal of Vocational and Career Education, 10*(1), 64–71. https://doi.org/10.15294/jvce.v10i1.30589

Satriawan, R., Sudarsono, B., Ghozali, F. A., & Pratama, W. (2025). Development and feasibility evaluation of electric vehicle learning kit (EV-KIT) based learning media for vocational high school students. *VANOS Journal of Mechanical Engineering Education, 10*(2), 272–286. https://doi.org/10.30870/vanos.v10i2.35488

Schaafstal, A., Schraagen, J. M., & van Berlo, M. (2000). Cognitive task analysis and innovation of training: The case of structured troubleshooting. *Human Factors, 42*(1), 75–86. https://doi.org/10.1518/001872000779656570

Siswanto, H., Tuwoso, & Suhartadi, S. (2024). Pengembangan trainer kendaraan listrik sederhana sebagai media pembelajaran pada mata pelajaran dasar otomotif di jurusan teknik otomotif [Development of a simple electric vehicle trainer as learning media in basic automotive subjects]. *Briliant: Jurnal Riset dan Konseptual, 9*(3), 716–728. https://doi.org/10.28926/briliant.v9i3.1763

Sugiyono. (2024). *Metode penelitian kuantitatif, kualitatif, dan R&D* [Quantitative, qualitative, and R&D research methods] (2nd ed.). Alfabeta. https://openlibrary.telkomuniversity.ac.id/home/catalog/id/231325/slug/metode-penelitian-kuantitatif-kualitatif-dan-r-d-edisi-2.html

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Taufiq, I., Setiawan, M. Y., Sugiarto, T., & Saputra, H. D. (2024). Pengaruh penerapan media pembelajaran CVT berbasis simulator terhadap hasil belajar siswa di SMK N 2 Payakumbuh [The effect of simulator-based CVT learning media on student learning outcomes at SMK N 2 Payakumbuh]. *JTPVI: Jurnal Teknologi dan Pendidikan Vokasi Indonesia, 2*(1), 119–130. https://doi.org/10.24036/jtpvi.v2i1.148

Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional development for training teachers of exceptional children: A sourcebook*. Leadership Training

Institute/Special Education, University of Minnesota. https://eric.ed.gov/?id=ED090725

Tohadi, T., Samani, M., & Susila, I. W. (2020). Development of the 4E-FTE engine simulator trainer to improve learning outcomes through problem based learning. *International Journal for Educational and Vocational Studies, 2*(8), 627–633. https://doi.org/10.29103/ijevs.v2i8.2692

Vickers, A. J., & Altman, D. G. (2001). Analysing controlled trials with baseline and follow up measurements. *BMJ, 323*(7321), 1123–1124. https://doi.org/10.1136/bmj.323.7321.1123

Wawan, & Arizal, H. (2026). Pengembangan trainer sistem penerangan berbasis simulator pada mata pelajaran kelistrikan otomotif di SMK Negeri 3 Buduran [Development of a simulator-based lighting system trainer for automotive electrical subjects at SMK Negeri 3 Buduran]. *Jurnal Pendidikan Teknik Mesin, 15*(1). https://ejournal.unesa.ac.id/index.php/jurnal-pendidikan-teknik-mesin/article/view/74169

Willson, V. L., & Putnam, R. R. (1982). A meta-analysis of pretest sensitization effects in experimental design. *American Educational Research Journal, 19*(2), 249–258. https://doi.org/10.3102/00028312019002249

Zacharia, Z. C., & Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. *Learning and Instruction, 21*(3), 317–331. https://doi.org/10.1016/j.learninstruc.2010.03.001

SUPPLEMENTARY MATERIAL

Supplementary Table S1. Indicator-level results of expert validation

Instrument and aspect	Items	Range of experts (%)	Mean (%)	Category
Material (3 experts)				
Concept accuracy	1–8	87.50–92.50	89.17	Very valid
Curriculum fit	9–11	86.67–93.33	88.89	Very valid
System representation	12–15	90.00–95.00	93.33	Very valid
Completeness of content	16–18	93.33–100	95.56	Very valid
Learning (fault) scenarios	19–20	90.00–100	93.33	Very valid
Media/hardware (3 experts)				
Hardware engineering	1–5	88.00–92.00	89.33	Very valid
Safety and construction	6–13	92.50–95.00	94.17	Very valid
Panel and readability	14–20	88.57–94.29	90.48	Very valid
System function	21–26	96.67–100	97.78	Very valid
OBD-II diagnosis	27–29	100	100.00	Very valid
Operation and maintenance	30–33	85.00–95.00	88.33	Very valid
Manual and documentation	34–37	85.00–95.00	90.00	Very valid
Learning/test instrument (2 experts)				
Competence fit	1–3	93.33	93.33	Very valid
Cognitive level	4–10	91.43–97.14	94.29	Very valid
Concept accuracy	11–14	95.00	95.00	Very valid
Clarity of wording	15–18	90.00	90.00	Very valid
Quality of options	19–22	90.00–95.00	92.50	Very valid
Time allocation	23	80.00	80.00	Valid