



DOI: <https://doi.org/10.38035/jmpis.v7i3>  
<https://creativecommons.org/licenses/by/4.0/>

## Evaluasi Program Pelatihan *Basic Technical Course* dengan Pendekatan Model Kirkpatrick di PT United Tractors Pandu Engineering

Dicky Dwi Saputra<sup>1\*</sup>, Nanang Alamsyah<sup>2</sup>, Yulizar Widiatama<sup>3</sup>

<sup>1</sup>UPN “Veteran” Jakarta, Jakarta, Indonesia, [2210312012@mahasiswa.upnvj.ac.id](mailto:2210312012@mahasiswa.upnvj.ac.id)

<sup>2</sup>UPN “Veteran” Jakarta, Jakarta, Indonesia, [nanang.alamsyah@upnvj.ac.id](mailto:nanang.alamsyah@upnvj.ac.id)

<sup>3</sup>UPN “Veteran” Jakarta, Jakarta, Indonesia, [yulizar.w@upnvj.ac.id](mailto:yulizar.w@upnvj.ac.id)

\*Corresponding Author: [2210312012@mahasiswa.upnvj.ac.id](mailto:2210312012@mahasiswa.upnvj.ac.id)

**Abstract:** *This study evaluates the effectiveness of the Basic Technical Course training program for Material Engineers at PT United Tractors Pandu Engineering using the Kirkpatrick evaluation model across three levels: Reaction (Level 1), Learning (Level 2), and Behavior (Level 3). The study employed a quantitative descriptive-inferential design with a saturated sample of 17 offline participants. Data were collected through Likert-scale questionnaires for Levels 1 and 3, and final test scores for Level 2. Results indicated that Level 1 achieved a grand mean of 4.699 (Very Satisfied); Level 2 demonstrated that participants' mean scores of 81.222 significantly exceeded the minimum competency standard of 75 ( $t = 6.000$ ,  $p < 0.001$ , Cohen's  $d = 1.455$ ); Level 3 yielded an overall mean of 2.758 (Good), with Application of Skills as the only dimension below threshold (mean = 2.447, Less Good), indicating a training transfer gap. The training program demonstrated strong effectiveness at the reaction and learning levels but requires structured post-training intervention to improve skill transfer to the workplace. This evaluation covers only Levels 1–3; Level 4 (Results) was not analyzed due to data accessibility and timeline constraints.*

**Keywords:** *Kirkpatrick Evaluation Model, Training Effectiveness, Basic Technical Course, Learning Transfer, Competency Assessment*

**Abstrak:** Penelitian ini mengevaluasi efektivitas program pelatihan *Basic Technical Course* untuk Material Engineer di PT United Tractors Pandu Engineering menggunakan model evaluasi Kirkpatrick pada tiga level: Reaksi (Level 1), Pembelajaran (Level 2), dan Perilaku (Level 3). Penelitian menggunakan pendekatan kuantitatif deskriptif-inferensial dengan sampel jenuh 17 peserta pelatihan offline. Data dikumpulkan melalui kuesioner skala Likert untuk Level 1 dan Level 3, serta nilai ujian akhir untuk Level 2. Hasil menunjukkan Level 1 mencapai grand mean 4,699 (Sangat Puas); Level 2 membuktikan rata-rata skor peserta sebesar 81,222 melebihi standar Kriteria Ketuntasan Minimal 75 secara signifikan ( $t = 6,000$ ;  $p < 0,001$ ; Cohen's  $d = 1,455$ ); Level 3 menghasilkan *mean* keseluruhan 2,758 (Baik), dengan dimensi Penerapan Keterampilan sebagai satu-satunya dimensi berkategori Kurang Baik ( $mean = 2,447$ ), mengindikasikan adanya kesenjangan transfer pelatihan. Program pelatihan terbukti

efektif pada level reaksi dan pembelajaran, namun memerlukan mekanisme intervensi pascapelatihan yang terstruktur guna meningkatkan transfer keterampilan ke tempat kerja. Evaluasi ini hanya mencakup Level 1–3; Level 4 (Hasil/*Results*) tidak dianalisis karena keterbatasan akses data kinerja organisasional dan rentang waktu observasi.

**Kata Kunci:** Model Evaluasi Kirkpatrick, Efektivitas Pelatihan, *Basic Technical Course*, Transfer Pembelajaran, Asesmen Kompetensi

---

## PENDAHULUAN

Investasi dalam pengembangan sumber daya manusia melalui pelatihan merupakan salah satu strategi kritis bagi perusahaan yang bergerak di sektor industri alat berat. PT United Tractors Pandu Engineering, sebagai anak perusahaan PT United Tractors Tbk yang berfokus pada rekayasa dan pemeliharaan alat berat, menyelenggarakan program *Basic Technical Course* sebagai program *onboarding* wajib bagi *New Employee Engineering*. Program ini mencakup enam materi teknis inti yang berlangsung selama dua minggu, meliputi *Engine System*, *Machine Element*, *Electrical System*, *Transmission System*, *Hydraulic and Pneumatic System*, serta *Track and Wheel Undercarriage*. Mengingat kompleksitas teknis pekerjaan di bidang ini, keberhasilan program pelatihan tidak cukup hanya diukur dari kepuasan peserta, melainkan harus dievaluasi secara sistematis dan multilevel untuk memastikan bahwa pembelajaran benar-benar terwujud dan diterapkan di tempat kerja.

Meskipun investasi pelatihan di industri manufaktur terus meningkat, sebagian besar organisasi hanya mengukur kepuasan peserta pada Level 1 (Reaksi) tanpa menindaklanjuti apakah pembelajaran benar-benar terjadi dan diterapkan di tempat kerja (Kirkpatrick & Kirkpatrick, 2006). Evaluasi yang mencakup Level 3 (Perilaku/transfer ke tempat kerja) masih sangat jarang dilakukan, khususnya dalam konteks pelatihan teknis industri alat berat di Indonesia. Kondisi ini menyebabkan keputusan tentang kelangsungan atau perbaikan program pelatihan tidak didasarkan pada data yang komprehensif (Phillips, 2002). Kesenjangan ini semakin kritis di lingkungan industri teknis karena kegagalan transfer keterampilan dapat berdampak langsung pada produktivitas, keselamatan kerja, dan efisiensi operasional (Grossman & Salas, 2011).

Kebaruan penelitian ini terletak pada penerapan model Kirkpatrick secara *multilevel* pada konteks pelatihan teknis industri alat berat di Indonesia, mengintegrasikan analisis kuantitatif parametrik yang jarang digunakan dalam evaluasi pelatihan skala kecil, serta mengidentifikasi secara spesifik dimensi perilaku kerja yang mengalami kesenjangan *transfer* pelatihan. Penelitian ini berbeda dari penelitian sebelumnya yang umumnya hanya mencakup satu atau dua level Kirkpatrick, atau menggunakan sampel organisasi besar sehingga temuan kurang dapat diterapkan pada perusahaan dengan batch pelatihan kecil (Tamkin, 2002). Pendekatan kuantitatif deskriptif-inferensial yang digunakan memungkinkan penarikan kesimpulan yang kuat meskipun dengan ukuran sampel terbatas, khususnya melalui penggunaan uji parametrik One-Sample T-Test yang tervalidasi secara statistik (Creswell & Creswell, 2018). Beberapa penelitian sebelumnya telah menerapkan model Kirkpatrick dalam evaluasi pelatihan teknis (Sitzmann dkk., 2006). Dalam meta-analisis terhadap 162 studi, ditemukan bahwa reaksi afektif peserta (kepuasan) hanya berkorelasi lemah dengan hasil pembelajaran ( $r = 0,09$ ), mengonfirmasi bahwa tingginya kepuasan peserta tidak serta-merta mencerminkan pencapaian kompetensi yang sesungguhnya (Noe dkk., 2018). Mengemukakan bahwa Level 3 (Behavior) adalah level yang paling sulit diukur, namun paling bernilai bagi organisasi karena langsung berkaitan dengan kinerja aktual (Holton, 2005). Menyatakan bahwa kesenjangan antara pembelajaran di kelas dan penerapan di tempat kerja merupakan tantangan universal dalam evaluasi pelatihan, yang terjadi akibat faktor lingkungan kerja dan motivasi

pascapelatihan. Praslova (2010) menggarisbawahi pentingnya relevansi konten pelatihan terhadap tugas kerja aktual sebagai prediktor utama *transfer*. Sementara itu, Gegenfurtner (Gegenfurtner, 2011) menemukan bahwa motivasi transfer memediasi hubungan antara kualitas program pelatihan dan perilaku kerja pascapelatihan, yang memperkuat pentingnya evaluasi Level 3 secara sistematis. Tujuan penelitian ini adalah mengevaluasi efektivitas program *Basic Technical Course* pada ketiga level Kirkpatrick tersebut dan memberikan rekomendasi berbasis data untuk perbaikan program secara berkelanjutan.

## METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif-inferensial, yaitu pendekatan yang menggabungkan statistik deskriptif untuk menggambarkan karakteristik data dengan statistik inferensial untuk menarik kesimpulan berbasis uji hipotesis (Creswell & Creswell, 2018). Pemilihan pendekatan ini didasarkan pada ketersediaan data numerik yang terstruktur dan ketiadaan data pre-test yang menjadi kendala umum dalam evaluasi pelatihan di lingkungan industri. Penelitian dilaksanakan di PT United Tractors Pandu Engineering, Cikarang, Jawa Barat, pada periode Oktober 2025 hingga Maret 2026.

Populasi penelitian adalah seluruh peserta pelatihan Basic Technical Course untuk Material Engineer yang mengikuti sesi pelatihan secara offline di Patria HO, berjumlah 17 orang. Penetapan populasi pada peserta offline didasarkan pada pertimbangan homogenitas kondisi belajar, akses langsung terhadap fasilitas workshop untuk praktik teknis, serta kemungkinan observasi perilaku di lingkungan kerja yang lebih valid pascapelatihan (Kirkpatrick & Kirkpatrick, 2006). Seluruh anggota populasi dijadikan subjek penelitian menggunakan teknik sampling jenuh (Prof. Dr. Sugiyono, t.t.) sehingga  $N = 17$  orang.

Instrumen penelitian terdiri dari tiga komponen. Pertama, kuesioner Level 1 (Reaksi) berisi 8 item pernyataan menggunakan skala Likert 5 poin (1 = Sangat Tidak Puas hingga 5 = Sangat Puas), mencakup dimensi: kejelasan tujuan, relevansi materi, kompetensi instruktur, kesesuaian durasi, keberagaman metode, fasilitasi partisipasi, ketersediaan fasilitas, dan kepuasan keseluruhan. Kedua, data Level 2 (Pembelajaran) berupa skor ujian akhir dari 5 mata pelajaran: Basic Machine Element, Basic Engine System, Basic Electrical System, Basic Transmission Torqflow, dan Basic Transmission Direct Drive. Nilai yang digunakan adalah nilai terbaik dari dua kesempatan ujian (Nilai Ujian 1 atau Nilai Ujian 2 apabila terdapat ujian remedial), dengan pertimbangan bahwa standar kompetensi bersifat kriteria minimum dan nilai terbaik mencerminkan potensi penguasaan kompetensi yang sesungguhnya. Konsekuensi dari pendekatan ini adalah bahwa tingkat kelulusan pada percobaan pertama, khususnya pada Basic Engine System (47,1%), tetap dilaporkan secara transparan sebagai indikator tingkat kesulitan relatif per kursus. Ketiga, kuesioner Level 3 (Perilaku) berisi 25 item pernyataan menggunakan skala Likert 4 poin (1 = Tidak Setuju hingga 4 = Sangat Setuju), dikelompokkan dalam 5 scope masing-masing 5 item: a) Decision-making and Action, b) Team Collaboration, c) Identify Equipment Control Issues, d) Application of Skills, dan e) Procedures and Communication Skills. Kuesioner Level 3 diisi oleh atasan langsung peserta guna meminimalkan bias self-report.

Analisis data mengikuti alur tiga tahap. Pada tahap pertama, dilakukan uji validitas instrumen Level 1 dan Level 3 menggunakan korelasi Pearson Product-Moment dengan  $r$ -tabel ( $n = 17$ ,  $df = 15$ ,  $\alpha = 0,05$ , two-tailed) = 0,482 sebagai nilai kritis. Uji reliabilitas dilakukan menggunakan Cronbach's Alpha dengan kategori: 0,70–0,79 = Acceptable, 0,80–0,89 = Good, dan  $\geq 0,90$  = Excellent (George & Mallery, 2003) Pada tahap kedua, analisis Level 1 menggunakan statistik deskriptif dengan kategori interpretasi: 4,21–5,00 = Sangat Puas, 3,41–4,20 = Puas, 2,61–3,40 = Cukup Puas, 1,81–2,60 = Tidak Puas, dan 1,00–1,80 = Sangat Tidak Puas. Analisis Level 2 dimulai dengan uji normalitas Shapiro-Wilk ( $n < 30$ ), dilanjutkan dengan One-Sample T-Test dengan  $H_0: \mu \leq 75$  dan  $H_1: \mu > 75$  pada  $\alpha = 0,05$  (one-tailed), serta

perhitungan effect size Cohen’s d. Analisis Level 3 menggunakan statistik deskriptif per scope dan per item dengan kategori: 3,26–4,00 = Sangat Baik, 2,51–3,25 = Baik, 1,76–2,50 = Kurang Baik, dan 1,00–1,75 = Tidak Baik. Pada tahap ketiga, dilakukan sintesis lintas level untuk mengidentifikasi kesenjangan transfer pelatihan. Seluruh analisis kuantitatif dilakukan menggunakan SPSS versi 26 pada tingkat kepercayaan 95%. Meskipun data Level 1 dan Level 3 bersumber dari responden yang berbeda (peserta dan atasan langsung), potensi common method bias tetap diminimalkan melalui pemisahan sumber data tersebut; tidak ditemukan indikasi bias metode umum yang serius dalam penelitian ini (Podsakoff dkk., 2003).

## HASIL DAN PEMBAHASAN

### Uji Validitas dan Reliabilitas Instrumen

Uji validitas instrumen Level 1 menghasilkan nilai r hitung untuk kedelapan item berkisar antara 0,708 hingga 0,938, seluruhnya melebihi r tabel (0,482), sehingga semua 8 item dinyatakan valid. Nilai Cronbach’s Alpha sebesar 0,9457 menunjukkan kategori Excellent ( $\alpha \geq 0,90$ ), mengindikasikan konsistensi internal instrumen yang sangat tinggi (Tavakol & Dennick, 2011). Uji validitas instrumen Level 3 menghasilkan nilai r hitung untuk 25 item berkisar antara 0,506 hingga 0,931, seluruhnya melebihi r tabel (0,482), sehingga semua 25 item dinyatakan valid. Cronbach’s Alpha keseluruhan sebesar 0,9681 (Excellent). Analisis reliabilitas per scope menunjukkan nilai alpha antara 0,883 (S5 – Procedures & Communication) hingga 0,963 (S2 – Team Collaboration), seluruhnya melebihi standar minimum 0,70. Ringkasan hasil uji reliabilitas disajikan pada Tabel 1.

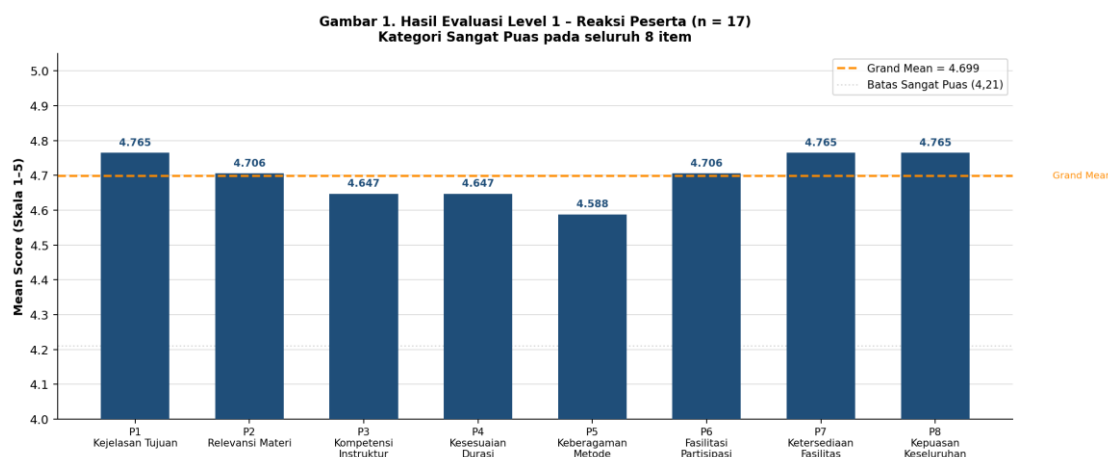
**Tabel 1. Ringkasan Uji Reliabilitas Instrumen Level 1 dan Level 3**

Instrumen	Jumlah Item	Cronbach's $\alpha$	Kategori	Keputusan
Level 1 – Kepuasan Peserta	8	0,9457	Excellent	Reliabel
Level 3 – S1: Decision-making & Action	5	0,9005	Excellent	Reliabel
Level 3 – S2: Team Collaboration	5	0,9626	Excellent	Reliabel
Level 3 – S3: Equipment Control Issues	5	0,9235	Excellent	Reliabel
Level 3 – S4: Application of Skills	5	0,8987	Good	Reliabel
Level 3 – S5: Procedures & Communication	5	0,8830	Good	Reliabel
<b>Level 3 – Keseluruhan 25 Item</b>	<b>25</b>	<b>0,9681</b>	<b>Excellent</b>	<b>Reliabel</b>

Sumber: Hasil pengolahan data primer, 2026

### Level 1 – Reaksi Peserta Terhadap Program Pelatihan

Analisis deskriptif terhadap 8 item kuesioner kepuasan peserta (n = 17) menunjukkan grand mean sebesar 4,699, yang berada pada kategori Sangat Puas (rentang 4,21–5,00). Distribusi individual menunjukkan 14 dari 17 responden (82,4%) berada pada kategori Sangat Puas, dan 3 responden (17,6%) pada kategori Puas. Tidak ada responden yang memberikan penilaian pada kategori Cukup Puas, Tidak Puas, maupun Sangat Tidak Puas. Hasil per item menunjukkan nilai mean berkisar antara 4,588 hingga 4,765. Rincian statistik deskriptif Level 1 disajikan pada Gambar 1.



**Gambar 1. Distribusi Mean Score Level**

Gambar 1 memperlihatkan distribusi yang relatif homogen dan konsisten tinggi pada seluruh 8 item kepuasan, dengan rentang mean antara 4,588 (P5 – Keberagaman Metode) hingga 4,765 (P1, P7, P8). Nilai terendah pada P5 mengindikasikan bahwa peserta menginginkan variasi metode pembelajaran yang lebih beragam, meskipun skor ini masih tergolong Sangat Puas. Secara keseluruhan, tidak ada item yang turun ke kategori di bawah Sangat Puas, yang menunjukkan kualitas desain dan pelaksanaan program yang konsisten. Statistik deskriptif lengkap per item disajikan pada Tabel 2 berikut.

**Tabel 2. Statistik Deskriptif Level 1 – Reaksi (n = 17)**

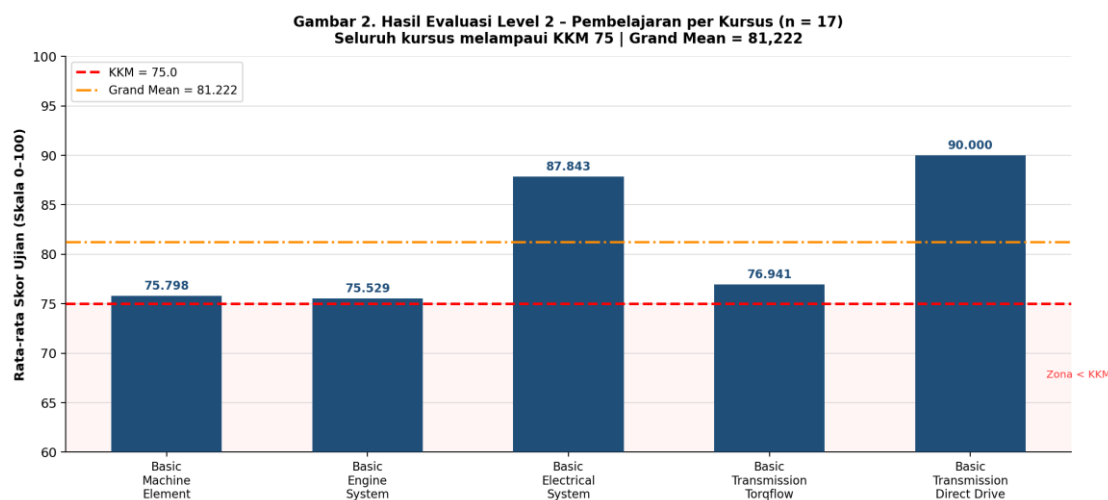
Item	Dimensi	Mean	SD	Min–Max	Kategori
P1	Kejelasan tujuan dan manfaat pelatihan	4,765	0,437	4–5	Sangat Puas
P2	Kemudahan dan relevansi materi	4,706	0,588	3–5	Sangat Puas
P3	Kompetensi instruktur	4,647	0,606	3–5	Sangat Puas
P4	Kesesuaian durasi pelatihan	4,647	0,606	3–5	Sangat Puas
P5	Keberagaman metode penyampaian	4,588	0,712	3–5	Sangat Puas
P6	Fasilitasi partisipasi aktif	4,706	0,588	3–5	Sangat Puas
P7	Ketersediaan fasilitas dan peralatan	4,765	0,437	4–5	Sangat Puas
P8	Kepuasan pengalaman pelatihan keseluruhan	4,765	0,562	3–5	Sangat Puas
<b>Grand Mean</b>		<b>4,699</b>	<b>0,563</b>		<b>Sangat Puas</b>

Sumber: Hasil pengolahan data primer, 2026

Temuan Level 1 ini sejalan dengan hasil penelitian (Alliger dkk., 1997) yang menyatakan bahwa reaksi positif peserta terhadap program pelatihan merupakan prasyarat penting bagi terjadinya pembelajaran yang efektif. Tingginya kepuasan terhadap kompetensi instruktur (P3: 4,647) dan ketersediaan fasilitas (P7: 4,765) mengindikasikan bahwa PT United Tractors Pandu Engineering telah menyediakan infrastruktur pelatihan yang memadai. Nilai P5 (Keberagaman Metode: 4,588) yang relatif lebih rendah meskipun masih dalam kategori Sangat Puas memberikan sinyal bahwa diversifikasi metode pembelajaran melalui penambahan simulasi atau pembelajaran berbasis masalah (*problem-based learning*) dapat meningkatkan keterlibatan peserta pada siklus pelatihan berikutnya (Sitzmann dkk., 2006). Hasil ini juga konsisten dengan temuan (Gegenfurtner, 2011) yang menunjukkan bahwa variasi metode instruksional berkontribusi positif terhadap motivasi transfer peserta pascapelatihan.

### Level 2 – Pencapaian Pembelajaran Peserta

Uji normalitas Shapiro-Wilk terhadap data grand mean skor ujian peserta (n = 17) menghasilkan statistik W = 0,959 dengan nilai signifikansi p = 0,606 > α = 0,05, sehingga asumsi normalitas terpenuhi untuk penggunaan uji parametrik (Shapiro & Wilk, 1965). One-Sample T-Test dilakukan untuk menguji H<sub>0</sub>: μ ≤ 75 melawan H<sub>1</sub>: μ > 75 dengan test value = 75 (Kriteria Ketuntasan Minimal). Hasil uji menunjukkan rata-rata grand mean peserta sebesar 81,222 (SD = 4,276), dengan nilai t-hitung = 6,000 yang melebihi t-tabel (df = 16, α = 0,05, one-tailed) = 1,746. Nilai signifikansi satu arah p < 0,001. Interval kepercayaan 95% [79,024; 83,421] sepenuhnya berada di atas nilai Kriteria Ketuntasan Minimal (75), memperkuat kesimpulan bahwa H<sub>0</sub> ditolak. Effect size Cohen’s d = 1,455 menunjukkan perbedaan yang sangat besar secara praktis, melampaui ambang batas large effect d > 0,80 (Cohen, 1988) Ringkasan hasil One-Sample T-Test disajikan pada Gambar 2.



Gambar 2. Rata-rata Skor Ujian per Kursus – Level 2

Gambar 2 memperlihatkan bahwa seluruh kursus berhasil melampaui ambang batas KKM 75, dengan rentang rata-rata skor antara 75,529 (Basic Engine System) hingga 90,000 (Basic Transmission Direct Drive). Variasi skor antar kursus ini mencerminkan perbedaan tingkat kesulitan materi; Basic Engine System dan Basic Machine Element mendekati batas KKM, sementara Basic Electrical System dan Basic Transmission Direct Drive menunjukkan penguasaan yang jauh lebih tinggi. Grand mean keseluruhan sebesar 81,222 secara statistik signifikan melebihi KKM, sebagaimana dikonfirmasi oleh uji One-Sample T-Test yang hasilnya dirinci pada Tabel 3 berikut.

Tabel 3. Hasil One-Sample T-Test Level 2 – Pembelajaran (N = 17, df = 16, Test Value = 75)

Statistik	N	Mean (SD)	t-hitung	p (one-tailed)	Cohen's d
<b>Grand Mean Skor Ujian</b>	<b>17</b>	<b>81,222 (4,276)</b>	<b>6,000*</b>	<b>&lt; 0,001</b>	<b>1,455</b>
95% CI Mean	–	[79,024; 83,421]	–	–	t-tabel = 1,746 (df = 16)

Sumber: Hasil pengolahan data primer, 2026

Ket.: Signifikan pada α = 0,05, one-tailed. df = N – 1 = 16.

Analisis per kursus (Tabel 4) menunjukkan variasi yang signifikan antar mata pelajaran. Basic Transmission Direct Drive memiliki mean tertinggi (90,000) dengan tingkat kelulusan pertama 100%, sementara Basic Engine System memiliki mean terendah (75,529) dengan tingkat kelulusan pertama hanya 47,1%. Fakta bahwa hanya 8 dari 17 peserta lulus Engine

System pada percobaan pertama mengindikasikan tingkat kesulitan relatif yang lebih tinggi pada materi ini, konsisten dengan temuan (Noe dkk., 2018) bahwa materi dengan konten konseptual tinggi cenderung lebih sulit dikuasai dibandingkan materi prosedural. Perlu dicatat bahwa nilai yang dilaporkan merupakan nilai terbaik dari dua kesempatan ujian; angka tingkat kelulusan pertama yang rendah pada Basic Engine System dan Basic Machine Element (58,8%) mencerminkan tantangan penguasaan konten yang sesungguhnya dan harus dipertimbangkan dalam evaluasi desain kurikulum, terlepas dari nilai akhir yang digunakan dalam analisis statistik.

**Tabel 4. Statistik Deskriptif Skor Ujian per Kursus (N = 17)**

Kursus	Mean	SD	Min <sup>1</sup>	Max <sup>1</sup>	Lulus KKM (≥75) Percobaan 1	Interpretasi
Basic Machine Element	75,798	7,429	54,29	85,71	10/17 (58,8%)	Tingkat kelulusan pertama rendah; materi perlu revisi alokasi waktu praktik
Basic Engine System	75,529	6,691	66,00	86,00	8/17 (47,1%)	Materi paling sulit; perlu redesain kurikulum dan penambahan durasi praktik
Basic Electrical System	87,843	8,328	70,00	100,00	16/17 (94,1%)	Efektivitas sangat baik; tingkat kelulusan pertama tertinggi kedua
Basic Transmission – Torqflow	76,941	5,750	60,00	84,00	14/17 (82,4%)	Efektivitas cukup baik; materi prosedural relatif mudah dikuasai
Basic Transmission – Direct Drive	90,000	8,101	75,00	100,00	17/17 (100,0%)	Efektivitas terbaik; seluruh peserta lulus pada percobaan pertama
<b>Grand Average<sup>2</sup></b>	<b>81,222</b>	<b>4,276</b>	<b>74,858</b>	<b>89,334</b>	<b>16/17 (94,1%)</b>	<b>Rata-rata efektivitas baik; terdapat variasi signifikan antar kursus</b>

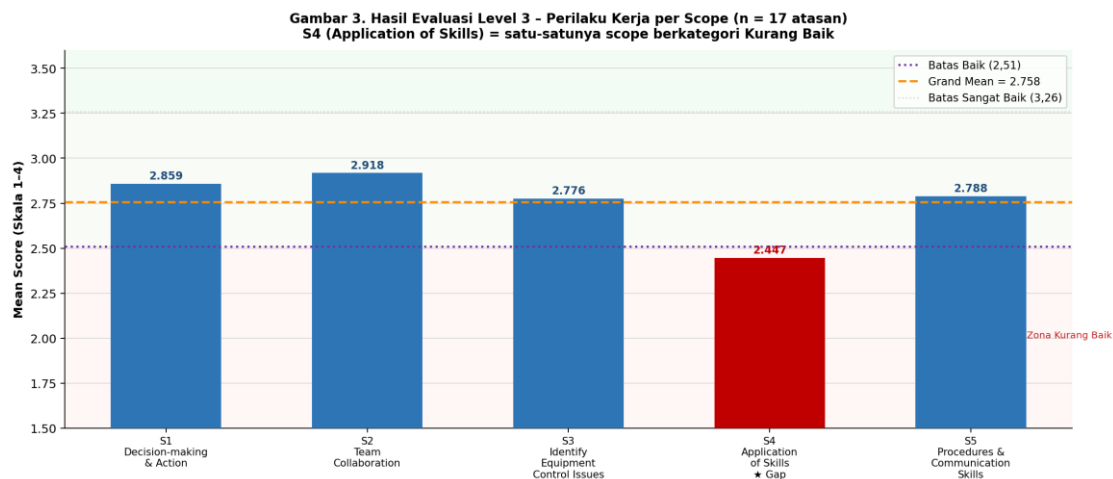
Sumber: Hasil pengolahan data primer, 2026

Ket.: <sup>1</sup>Min–Max per kursus = nilai ujian terendah dan tertinggi individual.

<sup>2</sup>Min–Max Grand Average = rata-rata grand mean peserta terendah dan tertinggi, bukan nilai ujian mentah individu.

### Level 3 – Perilaku Peserta di Tempat Kerja

Analisis Level 3 berdasarkan penilaian 17 atasan langsung peserta menggunakan instrumen 25 item skala Likert 4-poin menghasilkan grand mean keseluruhan sebesar 2,758, yang termasuk kategori Baik (rentang 2,51–3,25). Distribusi kategori individual menunjukkan 12 dari 17 atasan (70,6%) memberikan penilaian dalam kategori Baik, 4 atasan (23,5%) dalam kategori Kurang Baik, dan 1 atasan (5,9%) dalam kategori Sangat Baik. Tidak ada penilaian dalam kategori Tidak Baik. Analisis per scope disajikan pada Tabel 5 dan analisis per item untuk Scope S4 pada Gambar 3.



Gambar 3. Mean Score per Scope – Level 3 Perilaku Kerja (n=17 atasan)

Gambar 3 menunjukkan bahwa empat dari lima scope perilaku kerja (S1, S2, S3, S5) berada dalam kategori Baik dengan mean berkisar antara 2,776 hingga 2,918, yang mencerminkan adanya penerapan kompetensi yang memadai di tempat kerja. Satu-satunya scope yang berkategori Kurang Baik adalah S4 – Application of Skills (mean = 2,447), yang secara konsisten menunjukkan skor rendah pada semua item pengukurannya. Kesenjangan ini mengindikasikan bahwa peserta mampu memahami konsep teknis (Level 2) namun mengalami hambatan dalam mentransfer keterampilan tersebut ke kondisi kerja nyata. Rincian statistik deskriptif per scope dan per item disajikan pada Tabel 5 berikut.

Tabel 5. Statistik Deskriptif Level 3 per Scope (N = 17)

Scope	Grand Mean	SD	Range	Kategori
S1 – Decision-making and Action	2,859	0,423	2,00–3,60	Baik
S2 – Team Collaboration	2,918	0,534	1,80–4,00	Baik
S3 – Identify Equipment Control Issues	2,776	0,538	2,00–3,80	Baik
<b>S4 – Application of Skills</b>	<b>2,447</b>	<b>0,555</b>	<b>2,00–3,80</b>	<b>Kurang Baik</b>
S5 – Procedures and Communication Skills	2,788	0,450	2,00–3,80	Baik
<b>Grand Mean Keseluruhan</b>	<b>2,758</b>	<b>0,437</b>	<b>2,00–3,80</b>	<b>Baik</b>

Sumber: Hasil pengolahan data primer, 2026

Ket.: Cetak tebal menandai scope berkategori Kurang Baik.

Untuk memperoleh gambaran yang lebih terperinci mengenai scope S4 yang berada di bawah ambang kategori Baik, berikut disajikan analisis per item dalam Tabel 6.

Tabel 6. Analisis Per Item Scope S4 – Application of Skills (n = 17)

Item	Pernyataan	Mean	SD	Kategori
S4.1	Mampu mengoperasikan peralatan sesuai prosedur teknis yang diajarkan dalam pelatihan	2,412	0,507	Kurang Baik
S4.2	Mampu menerapkan teknik pemecahan masalah teknis di lapangan	2,412	0,515	Kurang Baik
S4.3	Mampu mengaplikasikan pengetahuan sistem mesin pada pekerjaan aktual	2,412	0,507	Kurang Baik

S4.4	Mampu mengidentifikasi dan menangani kerusakan komponen secara mandiri	2,412	0,515	Kurang Baik
S4.5	Menunjukkan peningkatan kecepatan dan akurasi kerja teknis pascapelatihan	2,576	0,507	Baik
<b>Rata-rata S4</b>		<b>2,447</b>	<b>0,555</b>	<b>Kurang Baik</b>

Sumber: Hasil pengolahan data primer, 2026

Ket.: Item S4.5 memiliki mean lebih tinggi (2,576 = Baik), sehingga rata-rata scope S4 sedikit di atas rata-rata empat item lainnya (2,412).

Temuan S4 (Application of Skills: 2,447– Kurang Baik) mengindikasikan bahwa penguasaan kompetensi di kelas belum cukup untuk menjamin penerapan di lapangan. Kesenjangan ini dapat dijelaskan oleh tiga faktor utama. Pertama, keterbatasan akses peralatan teknis di tempat kerja pascapelatihan; peserta yang baru bergabung umumnya belum mendapatkan penugasan langsung pada mesin atau sistem yang dipelajari, sehingga kesempatan untuk mempraktikkan keterampilan baru sangat terbatas (Ford dkk., 1992). Kedua, dukungan atasan dalam memfasilitasi transfer keterampilan berperan krusial namun sering kali tidak terstruktur; tanpa panduan atau coaching pascapelatihan yang eksplisit, karyawan baru cenderung mengandalkan prosedur standar yang lebih aman daripada menerapkan teknik baru yang lebih kompleks. Ketiga, budaya organisasi dan iklim kerja yang belum sepenuhnya mendukung eksplorasi teknis mandiri oleh karyawan baru dapat memperlambat transfer. (Baldwin & Ford, 1988) menegaskan bahwa faktor-faktor di luar kelas termasuk dukungan atasan, *opportunity to perform*, dan iklim organisasi merupakan prediktor transfer yang sama pentingnya dengan kualitas pelatihan itu sendiri. (Holton, 2005) juga menegaskan bahwa kesenjangan transfer paling sering terjadi pada keterampilan teknis yang membutuhkan peralatan dan kondisi kerja spesifik yang tidak selalu tersedia segera setelah pelatihan berakhir.

Sebaliknya, S2 – Team Collaboration memiliki mean tertinggi (2,918 – Baik), mengindikasikan bahwa program pelatihan berhasil membangun fondasi kolaborasi tim yang positif. Hal ini dapat dijelaskan oleh format pelatihan kelompok yang mendorong peserta berinteraksi aktif dalam sesi praktik dan *troubleshooting* bersama, sehingga keterampilan kolaborasi lebih mudah ditransfer ke konteks kerja (Noe dkk., 2018). (Deckop dkk., 2006) menemukan bahwa keterampilan interpersonal dan kolaborasi cenderung memiliki tingkat transfer yang lebih tinggi dibandingkan keterampilan teknis yang bergantung pada peralatan spesifik.

Perlu dicatat pula bahwa penilaian Level 3 bersumber dari atasan langsung peserta, bukan *self-report*. Meskipun penggunaan sumber eksternal ini memperkuat objektivitas data dibandingkan penilaian mandiri peserta, pendekatan ini tetap rentan terhadap *supervisor leniency bias* dan *halo effect*, yaitu kecenderungan atasan untuk memberikan penilaian yang lebih positif secara konsisten terhadap seluruh dimensi perilaku bawahan mereka (Podsakoff dkk., 2003). Sebagai implikasi metodologis, hasil Level 3 sebaiknya diperkuat melalui triangulasi dengan observasi langsung atau penilaian berbasis portofolio pekerjaan pada penelitian berikutnya guna meningkatkan validitas konstruk penilaian perilaku kerja (Grossman & Salas, 2011).

### Sintesis Lintas Level Kirkpatrick

Sintesis temuan ketiga level menghasilkan gambaran komprehensif tentang efektivitas program pelatihan. Pada Level 1, peserta menunjukkan reaksi sangat positif (grand mean 4,699), yang menjadi fondasi motivasional untuk terjadinya pembelajaran (Alliger dkk., 1997). Pada Level 2, pembelajaran terbukti efektif secara statistik, dengan 94,1% peserta melampaui

Kriteria Ketuntasan Minimal dan effect size yang besar ( $d = 1,455$ ). Pada Level 3, sebagian besar dimensi perilaku berada dalam kategori Baik, namun terdapat satu dimensi kritis (S4 – Application of Skills) yang memerlukan perhatian khusus.

Pola yang teridentifikasi menunjukkan bahwa program pelatihan berhasil pada tahap awal rantai nilai evaluasi, yaitu kepuasan dan pembelajaran, tetapi menghadapi tantangan pada tahap transfer ke tempat kerja, khususnya pada penerapan keterampilan teknis. Implikasi teoritis dari temuan ini mendukung model evaluasi Kirkpatrick yang menekankan perlunya intervensi pascapelatihan untuk memaksimalkan transfer pembelajaran ke perilaku kerja (Phillips, 2002). (Blume dkk., 2010) dalam meta-analisisnya mengonfirmasi bahwa intervensi pascapelatihan yang terstruktur secara konsisten menghasilkan peningkatan transfer yang signifikan, terutama bila melibatkan *supervisor coaching* dan *peer support*. (Grossman & Salas, 2011) lebih lanjut merekomendasikan bahwa program *on-the-job training* terstruktur pasca-classroom training merupakan intervensi paling efektif untuk mengatasi kesenjangan transfer pada keterampilan teknis industri.

Dalam konteks keterbatasan penelitian, perlu ditegaskan bahwa evaluasi Level 4 (Results) yang mengukur dampak pelatihan pada kinerja organisasi secara keseluruhan tidak termasuk dalam cakupan penelitian ini. Hal ini disebabkan oleh: a) Periode observasi pascapelatihan yang relatif singkat tidak mencukupi untuk mengidentifikasi perubahan kinerja organisasional yang terukur; b) Isolasi dampak pelatihan dari variabel-variabel konfundan seperti rotasi pekerjaan, perubahan manajemen, dan kondisi pasar memerlukan desain penelitian longitudinal yang lebih kompleks; serta c) Data kinerja organisasional di tingkat departemen bersifat konfidensial dan tidak sepenuhnya dapat diakses dalam konteks penelitian akademik. Keterbatasan ini disadari sepenuhnya dan menjadi agenda penelitian lanjutan yang direkomendasikan, sejalan dengan argumen (Tamkin, 2002) bahwa evaluasi Level 4 Kirkpatrick memerlukan komitmen organisasional dan sumber daya yang signifikan.

## KESIMPULAN

Evaluasi program Basic Technical Course menggunakan model Kirkpatrick tiga level menghasilkan temuan sebagai berikut. Level 1 mencapai grand mean 4,699 (Sangat Puas), menunjukkan desain program dan kualitas instruktur yang memenuhi ekspektasi peserta. Level 2 membuktikan rata-rata skor 81,222 secara statistis signifikan melampaui KKM 75 ( $t = 6,000$ ,  $p < 0,001$ ,  $d = 1,455$ ), mengkonfirmasi efektivitas transfer pengetahuan teknis. Level 3 menghasilkan grand mean 2,758 (Baik), namun dimensi Penerapan Keterampilan (S4, mean = 2,447) berkategori Kurang Baik, mengindikasikan adanya kesenjangan transfer dari kelas ke tempat kerja yang perlu ditangani.

Penelitian ini memiliki beberapa keterbatasan: (1) desain cross-sectional membatasi inferensi kausalitas; (2) ukuran sampel kecil ( $N = 17$ ) membatasi generalisasi; (3) Level 4 (Results) tidak dianalisis karena keterbatasan akses data kinerja organisasional dan rentang waktu yang singkat; dan (4) penilaian Level 3 oleh atasan berpotensi mengandung leniency bias (Podsakoff dkk., 2003). Keterbatasan-keterbatasan ini menjadi agenda penelitian lanjutan yang perlu diatasi.

Tiga rekomendasi praktis diajukan: (1) menyusun modul on-the-job coaching terstruktur 12 minggu pascapelatihan dengan checklist kompetensi per scope untuk menjembatani kesenjangan S4; (2) mengevaluasi ulang kurikulum Basic Engine System dengan penambahan durasi praktik mengingat tingkat kelulusan pertama hanya 47,1%; dan (3) mengintegrasikan simulasi berbasis komputer dan problem-based learning guna memperkuat aplikasi keterampilan teknis di tempat kerja. Secara akademis, penelitian lanjutan direkomendasikan menggunakan desain longitudinal dengan kelompok kontrol dan memperluas evaluasi ke Level 4 (Tamkin, 2002).

## REFERENSI

- Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A META-ANALYSIS OF THE RELATIONS AMONG TRAINING CRITERIA. *Personnel Psychology*, 50(2), 341–358. <https://doi.org/10.1111/j.1744-6570.1997.tb00911.x>
- Baldwin, T. T., & Ford, J. K. (1988). TRANSFER OF TRAINING: A REVIEW AND DIRECTIONS FOR FUTURE RESEARCH. *Personnel Psychology*, 41(1), 63–105. <https://doi.org/10.1111/j.1744-6570.1988.tb00632.x>
- 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/10.1177/0149206309352880>
- Cohen, J. (with Internet Archive). (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J.: L. Erlbaum Associates. [http://archive.org/details/statisticalpower0000cohe\\_j013](http://archive.org/details/statisticalpower0000cohe_j013)
- Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.
- Deckop, J. R., Konrad, A. M., Perlmutter, F. D., & Freely, J. L. (2006). The effect of human resource management practices on the job retention of former welfare clients. *Human Resource Management*, 45(4), 539–559. <https://doi.org/10.1002/hrm.20131>
- Ford, J. K., Quiñones, M. A., Sego, D. J., & Sorra, J. S. (1992). FACTORS AFFECTING THE OPPORTUNITY TO PERFORM TRAINED TASKS ON THE JOB. *Personnel Psychology*, 45(3), 511–527. <https://doi.org/10.1111/j.1744-6570.1992.tb00858.x>
- Gegenfurtner, A. (2011). Motivation and transfer in professional training: A meta-analysis of the moderating effects of knowledge type, instruction, and assessment conditions. *Educational Research Review*, 6(3), 153–168. <https://doi.org/10.1016/j.edurev.2011.04.001>
- George, D., & Mallery, P. (with Internet Archive). (2003). *SPSS for Windows step by step: A simple guide and reference, 11.0 update*. Boston: Allyn and Bacon. <http://archive.org/details/spssforwindowsst00geor>
- Grossman, R., & Salas, E. (2011). The transfer of training: What really matters. *International Journal of Training and Development*, 15(2), 103–120. <https://doi.org/10.1111/j.1468-2419.2011.00373.x>
- Holton, E. F. (2005). Holton's Evaluation Model: New Evidence and Construct Elaborations. *Advances in Developing Human Resources*, 7(1), 37–54. <https://doi.org/10.1177/1523422304272080>
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (with Internet Archive). (2006). *Evaluating training programs [electronic resource]: The four levels*. San Francisco, CA : Berrett-Koehler. [http://archive.org/details/Evaluating\\_Training\\_Programs\\_9781576757963](http://archive.org/details/Evaluating_Training_Programs_9781576757963)
- Noe, R. A., Hollenbeck, J. R., Gerhart, B., & Wright, P. M. (2018). *Human Resource Management: Gaining a Competitive Advantage*. McGraw-Hill Education.
- Phillips, J. J. (with Internet Archive). (2002). *How to measure training results: A practical guide to tracking the six key indicators*. New York: McGraw-Hill. <http://archive.org/details/howtomeasuretrai0000phil>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Prof. Dr. Sugiyono. (t.t.). *METODE PENELITIAN KUANTITATIF, KUALITATIF, DAN R&D*. Diambil 3 Mei 2026, dari <http://archive.org/details/buku-metode-penelitian-sugiyono>
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>

- Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006). THE COMPARATIVE EFFECTIVENESS OF WEB-BASED AND CLASSROOM INSTRUCTION: A META-ANALYSIS. *Personnel Psychology*, 59(3), 623–664. <https://doi.org/10.1111/j.1744-6570.2006.00049.x>
- Tamkin, P. (with Internet Archive). (2002). *Kirkpatrick and beyond: A review of training evaluation*. Brighton: Institute for Employment Studies. <http://archive.org/details/kirkpatrickbeyon0000tamk>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>