



Practical Demonstration Improves Assessment Performance in Farm Machinery Education: A Quasi-Experimental Study

*^{1,2}A.G.C. Saranga Amarasinghe & ¹Y.N.S. Wijewardana

¹Department of Engineering Technology, Faculty of Technological Studies, Uva Wellassa University of Sri Lanka, Passara Road, Badulla, Sri Lanka, 90000

²Department of Mechanical Engineering, College of Engineering (A), Andhra University, Visakhapatnam, Andhra Pradesh-530003, India

*beta17006@std.uwu.ac.lk

Received: 21 July 2026; Revised: 12 August 2026; Accepted: 15 September 2026; Available online: 10 Oct 2026

Abstract: The persistent emphasis on theory-led instruction in many engineering technology programmes raises a practical question: does direct, hands-on exposure to equipment measurably improve how well students understand and apply what they have already studied. This study evaluated the immediate effect of a practical demonstration on assessment performance among 75 final-year Engineering Technology undergraduates enrolled in the Farm Machinery Technology module at Uva Wellassa University of Sri Lanka. Using a two-group, order-manipulated (quasi-experimental) design, Group 1 ($n = 33$) completed a 25-item assessment immediately after a full-day practical demonstration of farm implements, tractor hitching, hydraulic operation and power take-off (PTO) applications, whereas Group 2 ($n = 42$) completed the same assessment before the demonstration. All students were advised to revise the same course material beforehand. Twenty closed-response items were scored against an answer key whose coding was validated against the authors' independent hand-tallies (exact agreement on 25 of 26 checkpoints). The practical-first group outperformed the quiz-first group (mean 19.06 ± 1.17 versus 18.24 ± 1.82 of 20; 95.3% versus 91.2%). Because scores were markedly non-normal and heteroscedastic, a Welch t-test and a Mann–Whitney U test were used; both indicated a statistically significant difference (Welch $t(70.5) = 2.37$, $p = 0.020$; $U = 884$, $p = 0.034$), with a medium effect size (Cohen's $d = 0.52$, 95% CI [0.11, 0.94]; Hedges' $g = 0.52$). The advantage was concentrated on items referring to implements that were physically present during the demonstration, and it disappeared for the one implement (the reciprocating mower) that was not available on site. After Holm and Benjamini–Hochberg correction across 20 items, no single item reached significance, indicating that the effect is real in aggregate but modest and diffuse rather than driven by isolated items. The findings support integrating structured practical demonstration alongside theoretical instruction, while cautioning that the benefit is bounded by physical access to the equipment and by an immediate-assessment design that does not test long-term retention.

Index Terms: Academic performance; Engineering technology education; Experiential learning; Farm machinery; Hands-on demonstration; Practical-based learning; Quasi-experimental design.

1. INTRODUCTION

Engineering technology education is, by its nature, an applied discipline: its graduates are expected not merely to recall principles but to recognise equipment, interpret its function and apply it competently in the field. Yet in many universities, and particularly in the Sri Lankan higher-education context, instruction remains weighted towards theory delivered in lecture halls, with comparatively limited structured contact with the machinery that students are expected to master [1], [2]. The concern is not that theoretical instruction is unimportant, but that theory absorbed in the abstract may not transfer readily to practical recognition and application when the learner has never handled, mounted or operated the equipment in question.

The idea that learning is deepened by concrete experience is long-standing. Kolb's experiential-learning cycle frames effective learning as a movement between concrete experience, reflective observation, abstract conceptualisation and active experimentation, implying that experience and abstraction are complementary rather than alternative routes to understanding [3]. Dewey had earlier argued that education divorced from experience risks becoming inert [4]. Later syntheses of how people learn and of active-learning practice in higher education reached a compatible view [5], [6]. More recently, a large body of quantitative evidence has accumulated in favour of active and experiential approaches: Freeman and colleagues' meta-analysis of 225 studies reported that active learning raised examination and concept-inventory scores by roughly 0.47 standard deviations and substantially reduced failure rates across science, technology, engineering and mathematics (STEM) disciplines [7], while Prince's influential review found broad, if uneven, support for the core elements of active, collaborative and problem-based learning in engineering [1]. Several mechanisms have been proposed for why hands-on exposure should help. Depth-of-processing and dual-coding accounts suggest that manipulating a physical object encodes information through multiple channels-visual, spatial and motor-producing a more durable and retrievable memory trace than a verbal description alone [8], [9], [10]. Retrieval-practice research further shows that actively using knowledge, rather than passively re-reading it, strengthens later performance [11], [12]. At the same time, the literature is not uniformly celebratory: cognitive-load theory and the guidance debate caution that unstructured discovery can overload novices, so that the benefit of practical work depends on how well it is structured and scaffolded [13], [14], [15]. Careful studies have also shown that students' feelings of learning can diverge from their actual learning, underlining the need to measure outcomes rather than rely on satisfaction [16], [17].

Empirical studies across disciplines echo this pattern. Practical-based learning has been reported as a vehicle for academic, technological and scientific education in engineering courses [18]. Structured practical work has been associated with improved achievement in secondary physics [19], practical research-based activity with stronger engagement and outcomes in medical physiology [20], simulation with better performance in undergraduate medical education [21], and laboratory-based and technology-active approaches with gains in vocational and electronics education [22], [23]. Experiential approaches have similarly been advocated in tourism and management education to bridge the theory-practice divide [24], and practical-based assessment has been linked to better problem-solving in basic science and technology [25], and the integration of theoretical and practical training has been examined even in instrumental music-performance education [26]. Not every study finds a strong effect-work on learning styles, for example, reports weak links between individual style preferences and attainment [27]-which reinforces the value of testing the effect directly in each new context rather than assuming it. Despite this breadth, three gaps motivate the present study. First, comparatively little of this quantitative evidence is drawn from engineering technology programmes in South Asian universities, and almost none from farm-machinery instruction specifically, where the objects of study are large, expensive implements that students rarely handle. Second, many prior studies report overall gains

without examining which kinds of items benefit, leaving open whether any advantage is general or is tied to specific, physically encountered content. Third, few studies pair their analysis with the methodological caution the design demands-distinguishing a genuine aggregate effect from artefacts of multiple testing, and separating immediate performance from long-term retention.

This study addresses those gaps using data from a full-day practical workshop conducted for final-year Engineering Technology undergraduates at Uva Wellassa University of Sri Lanka. The objectives were: (i) to quantify whether completing a 25-item assessment after a practical demonstration is associated with higher performance than completing it beforehand; (ii) to determine where in the assessment any advantage is concentrated; (iii) to estimate the magnitude and practical importance of the effect, distinct from its statistical significance; and (iv) to test whether any advantage is contingent on the physical availability of the equipment demonstrated. The corresponding research questions ask whether practical-first students score higher overall, whether the difference is statistically robust and of meaningful size, and whether item-level gains track the implements that students were actually able to see and touch.

2. LITERATURE REVIEW

2.1. Experiential and active learning as a theoretical foundation

Experiential-learning theory positions concrete experience as an engine of conceptual change rather than a mere illustration of it [3], [4]. Kolb and Kolb later extended this into the notion of learning spaces, arguing that the environment in which learning occurs shapes the depth of engagement [28]. Empirically, the strongest single piece of evidence remains Freeman and colleagues' meta-analysis, whose effect of about half a standard deviation is both statistically decisive and educationally meaningful across STEM fields [7]. Reviews specific to engineering reach compatible conclusions while stressing that the label "active learning" subsumes heterogeneous practices whose effectiveness varies [1], [29]. Frameworks such as ICAP-which ranks interactive over constructive, active and passive engagement-offer a way to understand why hands-on manipulation might outperform observation, which in turn outperforms listening [30].

2.2. Hands-on and laboratory learning in engineering

Within engineering, pedagogies of engagement and inductive methods have repeatedly been associated with improved conceptual understanding and skill development [29], [31], and the development of professional expertise is argued to depend on repeated, contextualised practice rather than declarative knowledge alone [2]. Michael's survey of the evidence base concludes that the case for active learning in the applied sciences is well supported, though he too warns against treating it as a panacea [32]. Laboratory and workshop settings are the archetypal site of such learning; Amida and colleagues, for instance, examined how different laboratory formats affected students' mastery of electrical principles and found format-dependent differences in performance [23].

2.3. Practical exposure, academic performance and retention

A recurring finding is that practical exposure improves not only immediate performance but the durability of learning, a claim grounded in retrieval-practice and depth-of-processing research [11], [12], [8]. Structured practical physics teaching has been shown to raise achievement relative to conventional instruction [19]; practical research-based activity has improved outcomes in physiology [20]; and simulation has enhanced both practical and written examination performance in medicine [21]. Multimodal and multimedia accounts explain these gains by the additional, non-verbal encoding that physical interaction provides [9], [10]. It is

important, however, to separate immediate post-exposure performance from genuinely long-term retention: the latter requires a delayed assessment, which many studies—including the present one—do not include.

2.4. Counter-evidence and boundary conditions

The literature also supplies useful caution. Cognitive-load theory warns that poorly structured practical tasks can overwhelm novices and depress learning [13], and the minimal-guidance debate argues that unguided discovery is frequently inferior to well-scaffolded instruction [14]. Problem-based learning is effective but demands careful design [15]. Studies of learning styles find little evidence that matching instruction to a preferred style improves attainment, undercutting simplistic “learners differ, so practise differs” arguments [27], [33]. Finally, the divergence between perceived and actual learning [17] means that positive student reactions to practical sessions cannot substitute for measured outcomes—the standard adopted here.

2.5. Positioning of the present study

Taken together, the evidence, read alongside research-based principles of how learning works [34], supports a qualified expectation: a well-structured practical demonstration should produce a measurable, though not necessarily large, improvement in the immediate application and recognition of previously studied material, with the size of the benefit likely to depend on how directly the assessment content maps onto what students physically encountered. The present study tests exactly this expectation in an under-represented setting—farm-machinery instruction within a South Asian engineering technology programme—and, unusually, uses the natural absence of one demonstrated implement as an internal check on whether the benefit is indeed tied to physical exposure.

3. METHODOLOGY

3.1. Research design

The study used a two-group, order-manipulated design that is best characterised as quasi-experimental. Both groups sat the identical 25-item assessment; the manipulated variable was whether the assessment was taken immediately after (Group 1) or before (Group 2) a common practical demonstration. Because students were assigned to the morning or afternoon cohort administratively rather than by random allocation, the design does not support strong causal inference and is reported here as quasi-experimental. The temporal structure of the day is shown in Fig. 1. As Fig. 1 illustrates, all 75 students had been advised to revise their notes and were then divided into two groups: in the first session (09:00–12:00), Group 1 ($n = 33$) attended the practical demonstration while Group 2 ($n = 42$) sat the 25-item assessment on prior knowledge only; in the second session (12:00–15:00), the order was reversed, so that Group 1 sat the assessment after practical exposure and Group 2 attended the demonstration.

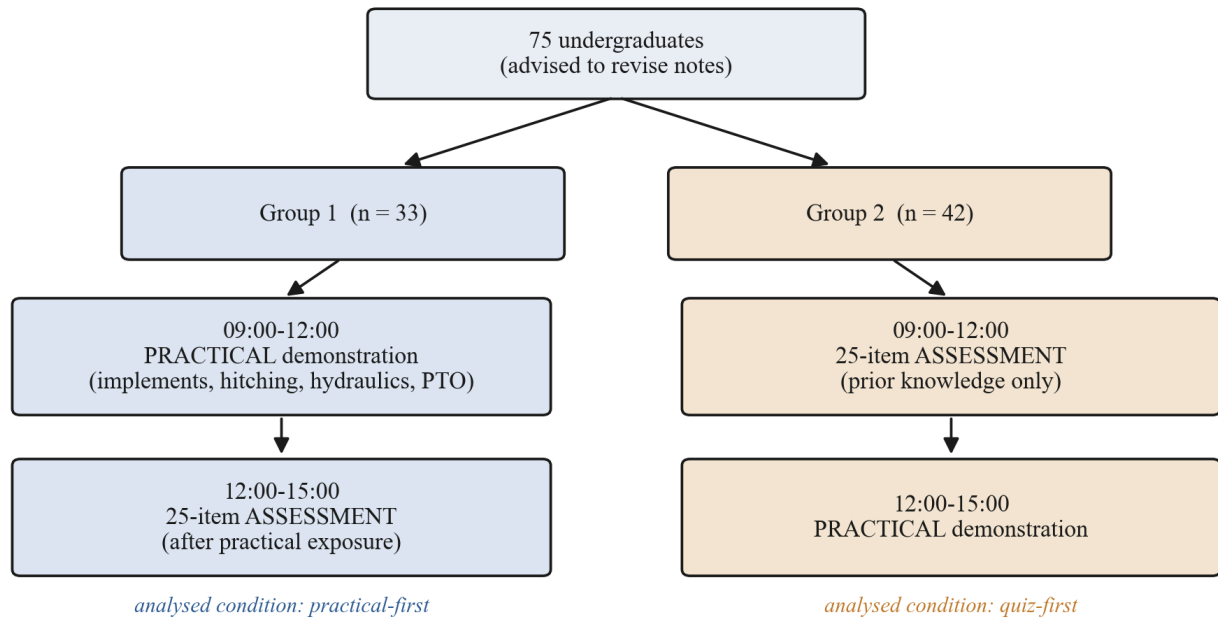
Experimental study design

Fig. 1. Study design and daily sequence. Both groups sat the same 25-item assessment; the analysed contrast is practical-first (Group 1) versus quiz-first (Group 2).

3.2. Study population and sample

Participants were 75 final-year undergraduates of the Department of Engineering Technology, Faculty of Technological Studies, Uva Wellassa University of Sri Lanka, all of whom had completed the Farm Machinery Technology 1 module. Group 1 comprised 33 respondents, and Group 2 comprised 42 respondents, giving a total of 75. All students were advised beforehand to revise the same lecture notes and reference materials, so that both groups approached the assessment from a comparable baseline of prior theoretical study.

3.3. Study context and instructional procedure

A full-day workshop was held on the university premises using the implements available on site. The demonstration covered land-clearing, primary-, secondary- and tertiary-tillage implements; the mounting of an implement to a tractor using the three-point hitch; the raising and lowering of implements through the hydraulic system; and the function of the power take-off (PTO). Students were able to observe and, in the case of hitching, physically handle the equipment listed in Table 1, which groups the implements by the field operation they serve, from land clearing through primary, secondary and tertiary cultivation to hitching and power transmission. One implement referred to in the assessment-the reciprocating mower-was not physically available at the university at the time of the workshop and was described only verbally; this fact is used later as an internal check.

Table 1. Farm implements and equipment used in the demonstration session, by operation.

Category	Implements/equipment demonstrated
Land clearing	Reciprocating mower (described only; not on site), rotary mower
Primary cultivation	Mould-board plough (MBP), disc plough
Secondary cultivation	Disc harrow, spike-tooth harrow, spring-tine harrow, cultivator
Tertiary operations	Subsoiler, rotavator, ridger
Hitching/power	Three-point hitch, one-point hitch, PTO shaft, tractor

3.4. Assessment instrument

The instrument was a 25-item online quiz delivered through an institutional Google Form requiring students to authenticate with their university email. It combined five open-response conceptual questions (Q1–Q5, e.g. definitions of primary and secondary cultivation), nine image-based implement-identification questions (Q7–Q15), seven closed-selection questions with named alternatives (Q6, Q17–Q21, Q25), and short-answer items on the PTO, hitching terminology and hydraulic function (Q16, Q22–Q24). The structure and the domain each item taps are summarised in Table 2. As Table 2 shows, twenty of the 25 items (Q6–Q25) required closed or short answers that could be scored objectively, whereas Q1–Q5 required open, essay-type responses.

Table 2. Structure of the 25-item assessment.

Items	Format	Primary domain assessed
Q1–Q5	Open response (essay)	Conceptual definitions (not machine-scored)
Q6, Q17–Q21, Q25	Closed selection	Application/implement selection
Q7–Q15	Image identification	Visual recognition of implements
Q16, Q22–Q24	Short answer	Function/terminology (PTO, hitch, hydraulics)

3.5. Variables and scoring

The independent variable was group (practical-first versus quiz-first). The primary dependent variable was the total closed-response score, computed over the twenty objectively scorable items (Q6–Q25). The five open-response items (Q1–Q5) were not scored because all students produced broadly correct definitions that differed only in richness; a binary correct/incorrect coding could not be defined objectively, and these items are therefore treated qualitatively and excluded from the numeric score. Each closed item was scored 1 for a correct response and 0 otherwise, using an answer key with normalisation for spelling and synonym variants (for example, “MBP”, “mould-board plough” and “mold bold plough” were treated as equivalent).

To guard against coding subjectivity, the automated scoring was validated against the authors’ own independent hand-tallies for every item for which a hand count existed: the automated and manual counts agreed exactly at 25 of 26 checkpoints, the single discrepancy being one Group 2 response to Q6 that the raw data record as “disc plough”.

3.6. Data processing and statistical analysis

Analysis was conducted in Python (pandas, NumPy, SciPy, statsmodels). Descriptive statistics were computed for each group. Because both score distributions were strongly left-skewed against a ceiling of 20 and failed the Shapiro–Wilk normality test, and because the variances were unequal by Levene’s test, group comparison relied on Welch’s unequal-variance t-test and the Mann–Whitney U test rather than the ordinary t-test. Effect sizes (Cohen’s d , Hedges’ g and the rank-biserial correlation) were reported with 95% confidence intervals; the interval for d was obtained by 10,000-replicate bootstrap resampling [35], [36]. At the item level, each group’s accuracy was compared using Fisher’s exact test with Haldane-corrected odds ratios; because twenty tests were performed, both Holm’s procedure and the Benjamini–Hochberg false-discovery-rate procedure were applied to control family-wise error and the false-discovery rate, respectively [37], [38]. Internal consistency of the twenty binary items was estimated using the Kuder–Richardson formula 20 (KR-20) [39], together with item difficulty and point-biserial discrimination. Statistical significance was set at $\alpha = 0.05$ (two-tailed).

The principal quantities are defined in equations (1)–(4). The group mean and sample standard deviation are

$$\bar{x} = \left(\frac{1}{n}\right) \sum x_i, \quad s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \quad (1)$$

the percentage accuracy for a group over the K scored items is

$$\text{Accuracy (\%)} = \left(\frac{\text{correct responses}}{\text{total responses}}\right) \times 10 \quad (2)$$

the standardised mean difference (Cohen’s d , with pooled standard deviation s_p) and its small-sample-corrected form (Hedges’ g) are

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_p}, \quad g = d \cdot \left[1 - \frac{3}{4(n_1 + n_2) - 9}\right] \quad (3)$$

and internal consistency is estimated as

$$\text{KR} - 20 = \left[\frac{K}{K - 1}\right] \cdot \left[1 - \frac{\sum p_i q_i}{\sigma_X^2}\right] \quad (4)$$

where p_i and q_i are the proportions answering item i correctly and incorrectly, and σ_X^2 is the variance of total scores.

3.7. Ethical considerations

Participation formed part of a scheduled academic workshop. Students authenticated with institutional email addresses; responses were used in aggregate for educational-evaluation purposes, and student index numbers are not reported in identifiable form. No formal ethics-committee approval number was recorded for this classroom-evaluation activity, and none is claimed here.

4. RESULTS AND DISCUSSION

4.1. Overall group performance

Across the twenty scored items, the practical-first group (Group 1) achieved a mean of 19.06 ± 1.17 (median 19.0; 95.3% accuracy), while the quiz-first group (Group 2) achieved 18.24 ± 1.82 (median 18.5; 91.2% accuracy). Descriptive statistics are given in Table 3, and the distributions in Fig. 2. In Fig. 2, both box plots are compressed against the maximum score of 20, and the Group 1 box is visibly narrower, reflecting its smaller interquartile range (1.00 versus 2.75 points; Table 3). Two features stand out. First, both groups performed very well: the assessment produced a pronounced ceiling effect, consistent with the fact that all students had been asked to revise the material in advance and were tested on content they had already studied. Second, Group 1's scores were both higher on average and less variable (coefficient of variation 6.1% versus 10.0%), suggesting that practical exposure not only lifted performance slightly but also compressed the lower tail of weaker responses.

Table 3. Descriptive statistics for total assessment score (out of 20) by group.

Statistic	Group 1 (practical-first)	Group 2 (quiz-first)
n	33	42
Mean	19.06	18.24
Median	19.0	18.5
Mode	19	20
Standard deviation	1.17	1.82
Variance	1.37	3.31
Minimum	15	13
Maximum	20	20
Range	5	7
Interquartile range	1.00	2.75
Coefficient of variation (%)	6.1	10.0
Skewness	-1.90	-1.25
Accuracy (%)	95.3	91.2

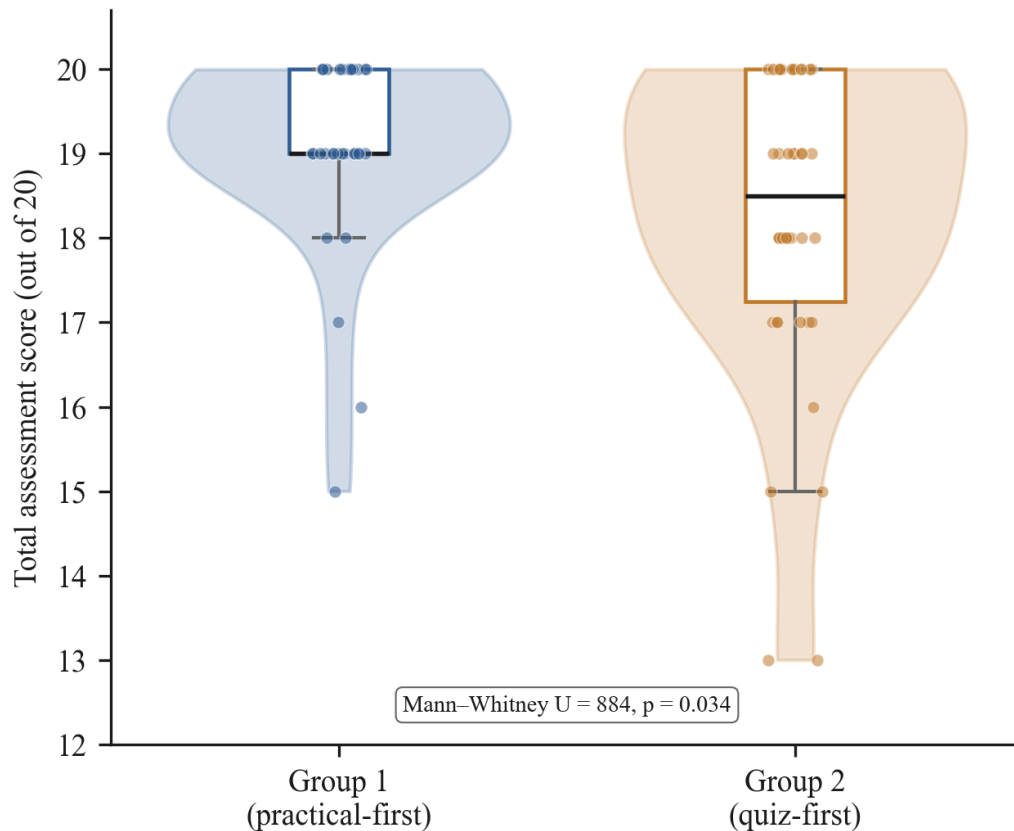
Distribution of total scores

Fig. 2. Distribution of total scores by group (box plots with individual observations; diamonds mark group means). Scores cluster against the ceiling of 20.

Because the distributions were far from normal (Shapiro–Wilk $p < 0.001$ for both groups) and the variances differed (Levene $p = 0.011$), inference was based on Welch’s t -test and the Mann–Whitney U test. Both agreed: the practical-first group scored significantly higher (Welch $t(70.5) = 2.37, p = 0.020$; Mann–Whitney $U = 884, p = 0.034$). The mean advantage was 0.82 points, or 4.1 percentage points, with a 95% confidence interval for the difference of [0.13, 1.51] points that excludes zero. The standardised effect was medium (Cohen’s $d = 0.52$, 95% CI [0.11, 0.94]; Hedges’ $g = 0.52$), and the common-language effect size of 0.64 indicates that a randomly chosen practical-first student would outscore a randomly chosen quiz-first student about 64% of the time. Inferential results and effect sizes are collected in Table 5 and Table 6.

4.2. Item-level performance

Disaggregating by item (Table 4, Fig. 3 and Fig. 4) clarifies where the aggregate advantage arises. The practical-first group led on the majority of items, but the margin was small for the many items that both groups answered almost perfectly, and concentrated on a few application items. The three largest practical-first advantages were the selection of the mould-board plough for obstacle-free land (Q18, +31.2 percentage points), the identification of weeders for shallow cultivation (Q21, +19.0 points) and the identification of the cultivator (Q15, +11.3 points)-all implements that were physically present and discussed during the demonstration.

Fig. 3 displays the accuracy of both groups on every scored item, ordered by the between-group difference,

and Fig. 4 isolates that difference in percentage points: bars to the right of zero favour Group 1, whereas the four bars to the left (Q14, Q16, Q20 and Q25) favour Group 2.

Table 4. Item-level accuracy, effect and difficulty (twenty scored items). Δ = Group 1 – Group 2 in percentage points; OR = Haldane-corrected odds ratio; p(F) = Fisher exact p; p(Holm) = Holm-adjusted p.

Item	Domain	G1 %	G2 %	Δ (pp)	OR	p(F)	p(Holm)
Q6	Application/selection	87.9	81.0	+6.9	1.62	0.532	1.000
Q7	Identification	100.0	100.0	+0.0	0.79	1.000	1.000
Q8	Identification	100.0	97.6	+2.4	2.42	1.000	1.000
Q9	Identification	100.0	92.9	+7.1	5.94	0.251	1.000
Q10	Identification	100.0	97.6	+2.4	2.42	1.000	1.000
Q11	Identification	97.0	92.9	+4.1	1.92	0.626	1.000
Q12	Identification	97.0	88.1	+8.9	3.18	0.220	1.000
Q13	Identification	100.0	95.2	+4.8	4.14	0.501	1.000
Q14	Identification	90.9	92.9	−1.9	0.77	1.000	1.000
Q15	Identification	97.0	85.7	+11.3	3.86	0.126	1.000
Q16	Conceptual	90.9	92.9	−1.9	0.77	1.000	1.000
Q17	Application/selection	100.0	100.0	+0.0	0.79	1.000	1.000
Q18	Application/selection	78.8	47.6	+31.2	3.88	0.008	0.160
Q19	Application/selection	100.0	95.2	+4.8	4.14	0.501	1.000
Q20	Application/selection	90.9	97.6	−6.7	0.31	0.314	1.000
Q21	Application/selection	100.0	81.0	+19.0	16.51	0.008	0.156
Q22	Conceptual	100.0	100.0	+0.0	0.79	1.000	1.000
Q23	Conceptual	100.0	100.0	+0.0	0.79	1.000	1.000
Q24	Conceptual	100.0	97.6	+2.4	2.42	1.000	1.000
Q25	Application/selection	75.8	88.1	−12.3	0.44	0.222	1.000

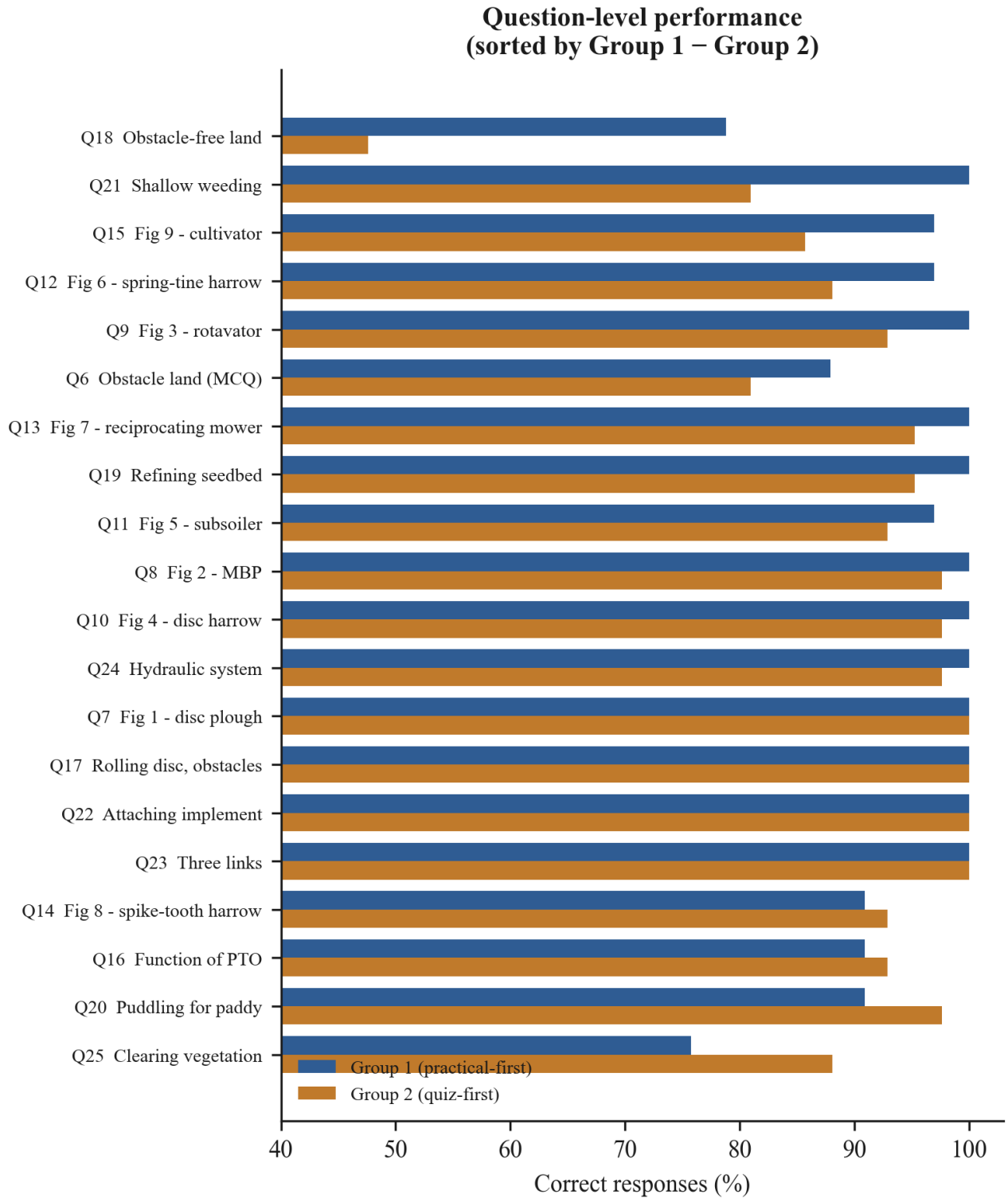


Fig. 3. Item-level accuracy for both groups, sorted by the Group 1 – Group 2 difference.

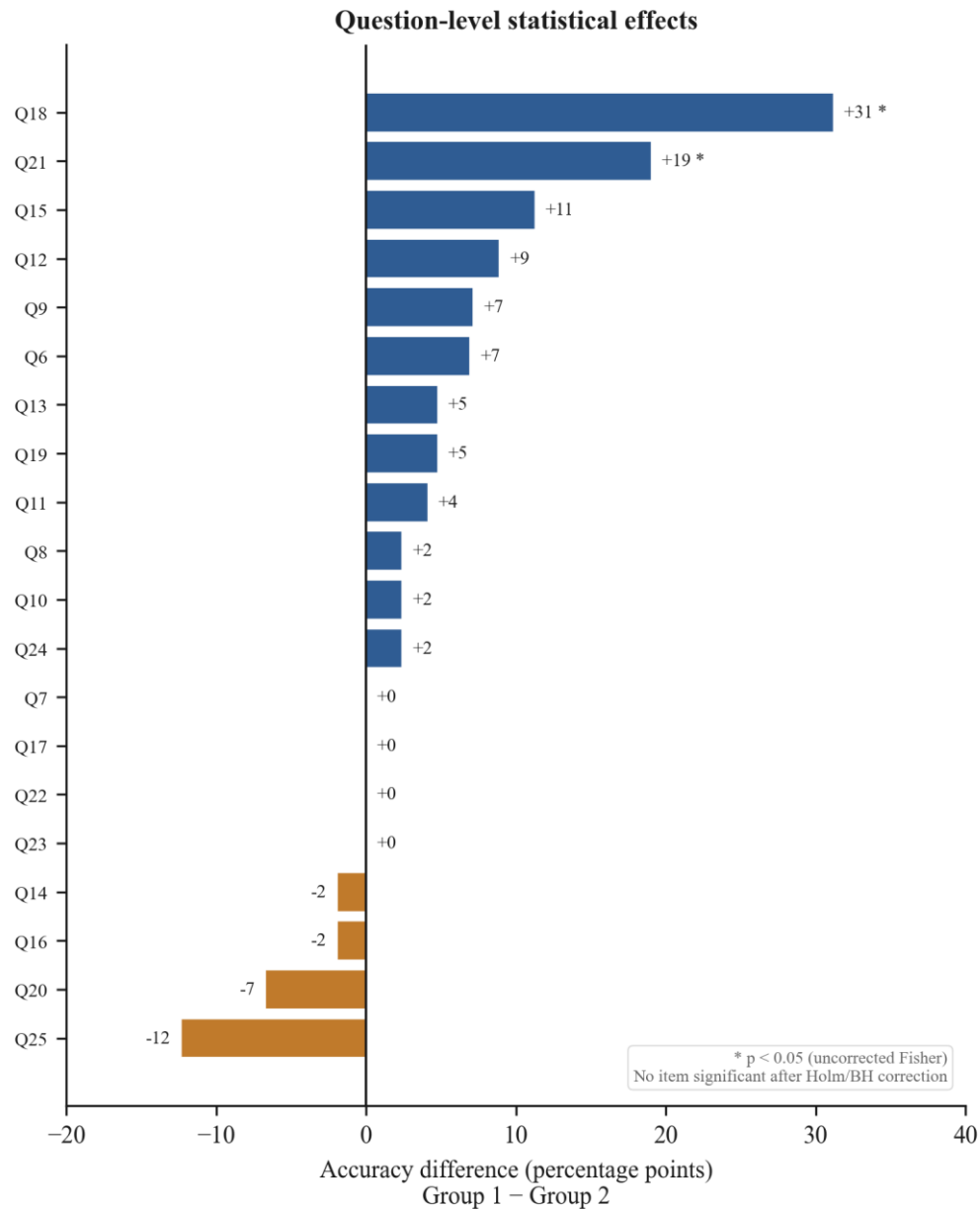


Fig. 4. Percentage-point difference in accuracy (Group 1 – Group 2) for each item. The reciprocating mower (Q25), which was not physically available, favours the quiz-first group.

Two cautions temper this picture. First, on the uncorrected Fisher tests, only Q18 and Q21 reached $p < 0.01$, and after correcting for twenty simultaneous comparisons, no individual item remained significant (smallest Holm-adjusted $p \approx 0.16$). The aggregate group difference is therefore real, but it is diffuse-spread thinly across many items-rather than attributable to a handful of decisive questions, and item-level claims should be read as descriptive. Second, the effect was not universal: the quiz-first group actually did slightly better on a few items, most notably the clearing-vegetation item (Q25, -12.3 points).

4.3. Physical availability as a boundary condition

The Q25 reversal is the study's most instructive detail. Q25 asks which implement clears vegetation; the intended answer is the mower-the one implement that was not physically present at the workshop and was described only verbally. On items whose equipment was available to see and handle, the practical-first group led or tied in most cases, and the few reversals (Q14, Q16 and Q20; -1.9 to -6.7 points) were small; on the

single item whose referent was absent, its advantage vanished and gave way to the largest reversal in the assessment (-12.3 points). Although this contrast rests on a single item and is not statistically significant on its own, it is consistent with the interpretation that the benefit of the demonstration was mediated by physical, multimodal encounter with the equipment rather than by any general motivational or order effect, and it accords with dual-coding and depth-of-processing accounts of why manipulation aids memory [8], [9], [10]. It also carries a concrete practical implication: the value of a demonstration is bounded by the equipment actually made available.

4.4. Reliability and item quality

Internal consistency of the twenty-item score was modest ($KR-20 = 0.59$ for the pooled sample; 0.44 and 0.63 within Groups 1 and 2). This is expected and, in context, unproblematic: the strong ceiling effect means that several items (for example, Q7, Q17, Q22 and Q23) were answered correctly by essentially everyone and therefore contribute no variance and no discrimination, mechanically depressing $KR-20$. Item-difficulty and point-biserial discrimination values (Fig. 5) confirm that most items were easy and weakly discriminating, so the assessment is better understood as a mastery check of previously studied content than as a finely discriminating test. The consequence for interpretation is that the true effect of practical exposure may be underestimated here, because a ceiling leaves little room for the stronger group to pull further ahead. In Fig. 5, most items lie at a proportion correct above 0.90, several fall below the conventional discrimination threshold of 0.20 (dashed line), and the items answered correctly by every student sit at zero discrimination; Q18 (proportion correct ≈ 0.61) is the only markedly harder item.

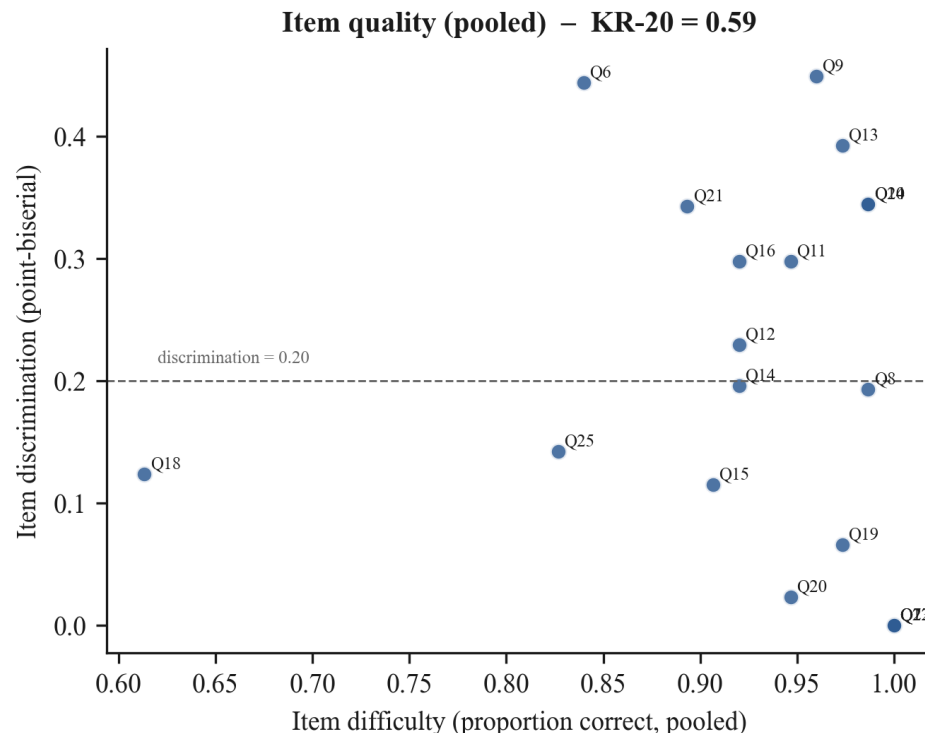


Fig. 5. Item difficulty against point-biserial discrimination (pooled sample). Most items are easy and weakly discriminating, reflecting the ceiling effect.

4.5. Inferential and effect-size summary

Table 5 Summarises the assumption checks that determined the choice of inferential tests. Both Shapiro–Wilk tests rejected normality ($W = 0.72$ and $W = 0.84$, both $p < 0.001$), and Levene’s test indicated unequal variances ($F = 6.79$, $p = 0.011$); Welch’s t-test and the Mann–Whitney U test were therefore used, and both detected a statistically significant group difference. Table 6 Reports the corresponding effect sizes: the mean difference of 0.82 points (95% CI [0.13, 1.51]) corresponds to 4.1 percentage points, Cohen’s d and Hedges’ g are both 0.52, the rank-biserial correlation is 0.28, and the common-language effect size of 0.64 expresses the same advantage as the probability that a practical-first student outscores a quiz-first student.

Table 5. Assumption checks and group-comparison tests for total score.

Test	Result
Shapiro–Wilk (Group 1)	$W = 0.72$, $p < 0.001$
Shapiro–Wilk (Group 2)	$W = 0.84$, $p < 0.001$
Levene (equal variance)	$F = 6.79$, $p = 0.011$
Welch t-test	$t(70.5) = 2.37$, $p = 0.020$
Mann–Whitney U	$U = 884$, $p = 0.034$

Table 6. Effect sizes for the group difference in total score (Group 1 – Group 2).

Measure	Value (95% CI)
Mean difference (points)	0.82 [0.13, 1.51]
Difference (percentage points)	4.1
Cohen’s d	0.52 [0.11, 0.94]
Hedges’ g	0.52
Rank-biserial correlation	0.28
Common-language effect size	0.64

4.6. Statistical versus practical significance

It is important to distinguish the two. The difference is statistically significant and of medium standardised size, yet in absolute terms it is a lift of roughly four percentage points from an already high base. For an educator, the practical reading is measured optimism: a single day of structured practical demonstration produced a detectable, medium-sized improvement in the immediate application of studied material and reduced the spread of weaker performances, but it did not transform outcomes, and much of its apparent modesty here is an artefact of the ceiling. The most defensible conclusion is that practical demonstration is a worthwhile complement to-not a replacement for-theoretical instruction, especially for content that can be tied to equipment students can actually handle.

4.7. Comparison with previous research and novel contribution

The direction and magnitude of the effect align with the wider literature. A medium effect ($d \approx 0.5$) is close to the 0.47-standard-deviation improvement reported in Freeman and colleagues' STEM meta-analysis [7] and consistent with the qualified support Prince describes for active learning in engineering [1]. The concentration of gains on application and identification, rather than on universally answered items, echoes findings that hands-on formats particularly aid tasks with a visual or procedural component [19], [21], [23]. The study's novel contributions are threefold: it provides quantitative evidence from an under-represented setting (farm-machinery instruction in a South Asian engineering technology programme); it links the item-level pattern of gains to the physical availability of specific equipment, using the naturally absent mower as an internal control; and it demonstrates methodological restraint by showing, through multiple-comparison correction, that a genuine aggregate effect need not imply significant differences at every item—a distinction often blurred in applied educational studies.

5. CONCLUSION

This study evaluated whether completing a 25-item assessment immediately after a practical demonstration improved the performance of final-year Engineering Technology undergraduates in Farm Machinery Technology relative to completing it beforehand. Students assessed after the demonstration scored significantly and meaningfully higher overall (95.3% versus 91.2%; Welch $p = 0.020$; Cohen's $d = 0.52$), with the advantage concentrated on items referring to equipment they were able to see and handle and absent for the one implement that was unavailable; no single item survived multiple-comparison correction, indicating a real but diffuse aggregate effect against a strong ceiling. The evidence supports integrating structured practical demonstration as a complement to theoretical instruction in engineering technology education, while recognising that the benefit is bounded by physical access to the equipment and that this immediate-assessment design speaks to understanding and application, not to long-term retention.

6. RECOMMENDATIONS

Directly supported by the results:

- Embed structured, guided practical demonstrations alongside theory in Farm Machinery Technology and comparable applied modules, since practical-first assessment was associated with higher and more consistent performance.
- Prioritise the physical availability of the specific implements to be assessed; the loss of advantage on the unavailable mower shows that verbal description is a weak substitute for hands-on exposure.
- Design assessments with greater ceiling headroom (harder application and transfer items) so that the effect of practical exposure can be measured more sensitively.

Broader, prudent extensions (future work):

- Adopt randomised allocation and, ideally, a counterbalanced crossover to strengthen causal inference beyond the present quasi-experimental design.
- Add a delayed re-assessment (for example at four to eight weeks) to test genuine retention rather than immediate performance.
- Replicate across multiple institutions, cohorts and applied domains to establish generalisability.

7. LIMITATIONS

Several limitations qualify these findings. The design was quasi-experimental: allocation to morning or afternoon cohorts was administrative rather than random, so unmeasured differences between groups cannot be excluded, and causal language is avoided accordingly. The study was conducted at a single university, in a single module, with a modest total sample ($n = 75$) and unequal group sizes. The assessment produced a strong ceiling effect and modest internal consistency ($KR-20 = 0.59$), which limits its discriminating power and probably attenuates the estimated effect. Five open-response items could not be objectively scored and were excluded from the numeric analysis. Most importantly, the assessment was administered immediately after the intervention, so the results speak to immediate understanding and application and provide no evidence about long-term retention, which the design was not built to measure. Finally, order effects and the novelty of a hands-on session cannot be fully separated from the practical content itself.

8. ACKNOWLEDGEMENT

The authors thank the staff and final-year Engineering Technology undergraduates of the Department of Engineering Technology, Uva Wellassa University of Sri Lanka, for facilitating and participating in the practical workshop.

9. REFERENCES

- [1]. M. Prince, "Does active learning work? A review of the research," *J. Eng. Educ.*, vol. 93, no. 3, pp. 223–231, 2004, doi: 10.1002/j.2168-9830.2004.tb00809.x.
- [2]. T. A. Litzinger, L. R. Lattuca, R. G. Hadgraft, and W. C. Newstetter, "Engineering education and the development of expertise," *J. Eng. Educ.*, vol. 100, no. 1, pp. 123–150, 2011, doi: 10.1002/j.2168-9830.2011.tb00006.x.
- [3]. D. A. Kolb, *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1984.
- [4]. J. Dewey, *Experience and Education*. New York, NY, USA: Kappa Delta Pi, 1938.
- [5]. C. C. Bonwell and J. A. Eison, *Active Learning: Creating Excitement in the Classroom*, ASHE-ERIC Higher Education Report No. 1. Washington, DC, USA: George Washington Univ., 1991.
- [6]. National Research Council, *How People Learn: Brain, Mind, Experience, and School*, expanded ed. Washington, DC, USA: National Academies Press, 2000.
- [7]. S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, "Active learning increases student performance in science, engineering, and mathematics," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, no. 23, pp. 8410–8415, 2014, doi: 10.1073/pnas.1319030111.
- [8]. F. I. M. Craik and R. S. Lockhart, "Levels of processing: A framework for memory research," *J. Verbal Learn. Verbal Behav.*, vol. 11, no. 6, pp. 671–684, 1972, doi: 10.1016/S0022-5371(72)80001-X.
- [9]. A. Paivio, "Dual coding theory: Retrospect and current status," *Can. J. Psychol.*, vol. 45, no. 3, pp. 255–287, 1991, doi: 10.1037/h0084295.
- [10]. R. E. Mayer, "Multimedia learning," *Psychol. Learn. Motiv.*, vol. 41, pp. 85–139, 2002, doi: 10.1016/S0079-7421(02)80005-6.
- [11]. H. L. Roediger and J. D. Karpicke, "Test-enhanced learning: Taking memory tests improves long-term retention," *Psychol. Sci.*, vol. 17, no. 3, pp. 249–255, 2006, doi: 10.1111/j.1467-9280.2006.01693.x.
- [12]. J. D. Karpicke and H. L. Roediger, "The critical importance of retrieval for learning," *Science*, vol. 319, no. 5865, pp. 966–968, 2008, doi: 10.1126/science.1152408.

- [13]. J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cogn. Sci.*, vol. 12, no. 2, pp. 257–285, 1988, doi: 10.1207/s15516709cog1202_4.
- [14]. P. A. Kirschner, J. Sweller, and R. E. Clark, "Why minimal guidance during instruction does not work," *Educ. Psychol.*, vol. 41, no. 2, pp. 75–86, 2006, doi: 10.1207/s15326985ep4102_1.
- [15]. C. E. Hmelo-Silver, "Problem-based learning: What and how do students learn?," *Educ. Psychol. Rev.*, vol. 16, no. 3, pp. 235–266, 2004, doi: 10.1023/B:EDPR.0000034022.16470.f3.
- [16]. L. Deslauriers, E. Schelew, and C. Wieman, "Improved learning in a large-enrollment physics class," *Science*, vol. 332, no. 6031, pp. 862–864, 2011, doi: 10.1126/science.1201783.
- [17]. L. Deslauriers, L. S. McCarty, K. Miller, K. Callaghan, and G. Kestin, "Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 116, no. 39, pp. 19251–19257, 2019, doi: 10.1073/pnas.1821936116.
- [18]. J. F. Zamberlan, G. E. Guimarães, G. C. Masutti, and R. F. dos S. Salazar, "Practical Based Learning (PBL) for academic, technological and scientific education in engineering courses – case study," *Int. J. Adv. Eng. Res. Sci.*, vol. 5, no. 6, pp. 131–134, 2018, doi: 10.22161/ijaers.5.6.22.
- [19]. M. A. Musasia, P. Thomas, and W. Sakwa, "Physics practical work and its influence on students' academic achievement," *J. Educ. Pract.*, vol. 7, no. 28, pp. 129–134, 2016.
- [20]. S. Ewida, "Assessment of a practical research-based activity program as a tool for teaching medical physiology to undergraduates," *J. Afr. Assoc. Physiol. Sci.*, vol. 2, no. 2, pp. 78–86, 2014.
- [21]. P. J. Morgan, D. Cleave-Hogg, S. DeSousa, and J. Lam-McCulloch, "Applying theory to practice in undergraduate education using high fidelity simulation," *Med. Teach.*, vol. 28, no. 1, pp. e10–e15, 2006, doi: 10.1080/01421590600568488.
- [22]. N. F. Hassan, S. Puteh, and R. Buhari, "Student understanding through the application of technology enabled active learning in practical training," *Procedia Soc. Behav. Sci.*, vol. 204, pp. 318–325, 2015, doi: 10.1016/j.sbspro.2015.08.158.
- [23]. A. Amida, I. Chang, and D. Yearwood, "Designing a practical lab-based assessment: a case study," *J. Eng. Des. Technol.*, vol. 18, no. 3, pp. 567–581, 2020, doi: 10.1108/JEDT-08-2019-0194.
- [24]. L. Ruhanen, "Bridging the divide between theory and practice: Experiential learning approaches for tourism and hospitality management education," *J. Teach. Travel Tour.*, vol. 5, no. 4, pp. 33–51, 2005, doi: 10.1300/J172v05n04_03.
- [25]. N. A. Omilani, S. A. Akinyele, T. S. Durowoju, and E. I. Obideyi, "The effect of the assessment of practical-based work on pupils' problem solving and achievement in basic science and technology," *Education 3-13*, vol. 47, no. 6, pp. 760–772, 2019, doi: 10.1080/03004279.2018.1534874.
- [26]. V. O. Malaschenko, M. A. Antonova, G. L. Knyazeva, I. A. Belokon, and B. A. Pechersky, "Theoretical and practical aspects of integrating the computer and music technologies into the instrumental performance training of the pedagogical university students," *SHS Web Conf.*, vol. 79, art. 01014, 2020, doi: 10.1051/shsconf/20207901014.
- [27]. T. Wilkinson, M. Boohan, and M. Stevenson, "Does learning style influence academic performance in different forms of assessment?," *J. Anat.*, vol. 224, no. 3, pp. 304–308, 2014, doi: 10.1111/joa.12126.
- [28]. A. Y. Kolb and D. A. Kolb, "Learning styles and learning spaces: Enhancing experiential learning in higher education," *Acad. Manag. Learn. Educ.*, vol. 4, no. 2, pp. 193–212, 2005, doi: 10.5465/amle.2005.17268566.
- [29]. M. J. Prince and R. M. Felder, "Inductive teaching and learning methods: Definitions, comparisons, and research bases," *J. Eng. Educ.*, vol. 95, no. 2, pp. 123–138, 2006, doi: 10.1002/j.2168-9830.2006.tb00884.x.
- [30]. M. T. H. Chi and R. Wylie, "The ICAP framework: Linking cognitive engagement to active learning outcomes," *Educ. Psychol.*, vol. 49, no. 4, pp. 219–243, 2014, doi: 10.1080/00461520.2014.965823.
- [31]. K. A. Smith, S. D. Sheppard, D. W. Johnson, and R. T. Johnson, "Pedagogies of engagement: Classroom-based practices," *J. Eng. Educ.*, vol. 94, no. 1, pp. 87–101, 2005, doi: 10.1002/j.2168-9830.2005.tb00831.x.
- [32]. J. Michael, "Where's the evidence that active learning works?," *Adv. Physiol. Educ.*, vol. 30, no. 4, pp. 159–167, 2006, doi: 10.1152/advan.00053.2006.

- [33]. R. M. Felder and L. K. Silverman, "Learning and teaching styles in engineering education," *Eng. Educ.*, vol. 78, no. 7, pp. 674–681, 1988.
- [34]. S. A. Ambrose, M. W. Bridges, M. DiPietro, M. C. Lovett, and M. K. Norman, *How Learning Works: Seven Research-Based Principles for Smart Teaching*. San Francisco, CA, USA: Jossey-Bass, 2010.
- [35]. J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ, USA: Lawrence Erlbaum, 1988.
- [36]. L. V. Hedges, "Distribution theory for Glass's estimator of effect size and related estimators," *J. Educ. Stat.*, vol. 6, no. 2, pp. 107–128, 1981, doi: 10.3102/10769986006002107.
- [37]. S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Stat.*, vol. 6, no. 2, pp. 65–70, 1979.
- [38]. Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. R. Stat. Soc. B*, vol. 57, no. 1, pp. 289–300, 1995, doi: 10.1111/j.2517-6161.1995.tb02031.x.
- [39]. G. F. Kuder and M. W. Richardson, "The theory of the estimation of test reliability," *Psychometrika*, vol. 2, no. 3, pp. 151–160, 1937, doi: 10.1007/BF02288391.