EFFECTIVENESS OF CTL- AND PROBLEM-BASED LEARNING WORKSHEETS IN ENHANCING MATHEMATICAL PROBLEM-SOLVING ABILITY AMONG STUDENTS OF SMK NEGERI 1 SEI SUKA ON SYSTEMS OF LINEAR EQUATIONS

Mardhotilla Utami [1], Tua Halomoan Harahap[2]
[1]Master of Mathematics Education Program, Graduate School, Universitas Muhammadiyah Sumatera Utara, Indonesia
[2]SMK Negeri 1 Sei Suka, Batu Bara Regency, North Sumatra, Indonesia
utamiyasir@gmail.com, tuahomoan@umsu.ac.id

ABSTRACT

This study evaluated the methodological coherence of Contextual Teaching and Learning and Problem-Based Learning worksheets designed to improve mathematical problem-solving ability in systems of linear equations. A non-equivalent control-group case-study design was modelled using 72 synthetic student records, with 36 records in an Office Management and Business Services condition and 36 in an Accounting and Institutional Finance condition. The six-unit worksheets incorporated vocational mathematical modelling scaffolding, progressive fading, model validation, and transfer/problem-posing activities. Five open-ended items were scored using four Polya indicators. Analysis of covariance adjusted posttest performance for pretest scores, while raw change, normalized gain, and robustness checks supported the main analysis. The adjusted posttest difference was 10.410 points (95% CI [5.947, 14.872]), $F(1,69)=21.656$, $p<.001$, partial eta-squared=.239, with adjusted d=1.097. Simulated reliability, content-validity, and fidelity evidence was also coherent. Because all key data were simulated and only one class/program was modelled per condition, the results did not establish causal effectiveness in real students.

Keyword : student worksheets, contextual teaching and learning, problem-based learning, mathematical problem solving, linear equations, vocational education, ANCOVA

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:
Mardhotilla Utami
Master of Mathematics Education Program, Graduate School,
Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia.
Jalan Kapten Mukhtar Basri No 3 Medan 20238, Indonesia.
Email: utamiyasir@gmail.com

1. INTRODUCTION

Mathematical problem-solving ability is a fundamental competency because it requires students not merely to execute algorithms, but also to understand situations, select appropriate representations, construct mathematical models, interpret solutions, and evaluate the reasonableness of results. Polya (1973) proposed four core processes - understanding the problem, devising a plan, carrying out the plan, and looking back - which remain relevant as a cognitive framework for assessing mathematical problem solving. From the perspective of contemporary mathematical literacy, these processes are closely related to the ability to formulate real-world situations mathematically, employ mathematical concepts and procedures, and interpret and evaluate results within context [9].

In vocational education, these demands are more specific. Mathematics should be connected to meaningful workplace situations without requiring hidden vocational knowledge as a prerequisite for solving the tasks. CTL provides principles for linking concepts with students' contexts and experiences [7], whereas PBL positions problems as the driving force for inquiry, collaboration, presentation, and evaluation. Barrows [2] described PBL as a family of problem-centered approaches, while the CTL/REACT framework emphasizes relating, experiencing, applying, cooperating, and transferring [4]. In this study, these approaches are not proposed as a new theory; rather, they serve as pedagogical foundations for the design of the student worksheets.

The current state of the literature suggests a need for a more sharply articulated instructional design than simply combining multiple pedagogical models. The meta-analysis by Salsabila and Asih [12] indicated that PBL is generally positively associated with students' mathematical problem-solving ability; therefore, the use of PBL alone is insufficient to constitute novelty. In the vocational-school context, Rendrayana et al. [11] reported that context-based PBL may improve critical thinking, problem solving, conceptual understanding, and mathematical literacy, while also identifying challenges related to teacher readiness, limited availability of workplace-contextualized instructional materials, and time constraints.

These findings support the need for learning materials that are not merely contextual, but that also incorporate explicit modelling structures and scaffolding.

Recent research on metacognition and scaffolding further supports the gradual withdrawal of instructional assistance. Hidayat et al. [6] reported positive effects of metacognitive instruction on mathematics achievement and metacognitive skills. Thomann and Deutscher [13] showed that prompts can support learning, although their effectiveness is highly dependent on design and implementation. At the final stage, problem posing becomes important for examining transfer and learner agency; the meta-analysis by Zhang et al. [15] reported positive effects of problem-posing interventions. Accordingly, the worksheets in this case study were designed to include progressive fading, deliberate error/model-validation tasks, and transfer/problem-posing activities, rather than merely presenting word problems labelled as vocational contexts.

The research gap addressed in this study is the lack of a structured connection among vocational context, mathematical modelling, metacognitive scaffolding, and the assessment of mathematical problem solving within a single auditable evidence chain. The novelty of this case study lies in its vocational mathematical modelling scaffolding. Students move from a real situation through assumptions and simplification, formulation, solving/employing mathematics, interpretation, validation/revision, and finally transfer/problem posing. CTL/REACT functions primarily as a mechanism for contextualization and transfer; PBL provides the social-pedagogical structure; Polya's framework provides the cognitive structure and scoring basis; the modelling cycle bridges the real situation and its mathematical representation; and metacognitive scaffolding provides support that is progressively reduced over time.

The primary objective of this study is to evaluate the coherence of a quantitative design for assessing the effectiveness of CTL- and PBL-based student worksheets in improving mathematical problem-solving ability in systems of linear equations. Specifically, this case study aims to: (1) compare posttest scores adjusted for pretest performance using ANCOVA; (2) evaluate the consistency of results through raw change, N-Gain, and sensitivity analyses; (3) explore changes across the four Polya indicators; and (4) audit the consistency of the instrument, content validity, internal reliability, response-process evidence, and implementation fidelity as components of the overall evidence chain.

2. RESEARCH METHOD

A. Study Design and Participants

This manuscript is a methodological/academic case study employing a quantitative approach. The case study models a non-equivalent control-group design with a pretest-posttest structure [3]. The independent variable is the use of CTL- and PBL-based student worksheets incorporating vocational mathematical modelling scaffolding in the experimental condition. The dependent variable is mathematical problem-solving ability in systems of linear equations, assessed using a rubric based on four Polya indicators. All main datasets, pilot data, validator ratings, and fidelity data reported in this article are simulated/reconstructed and are used as demonstration evidence for evaluating methodological coherence.

Within the proposed research design, the population comprises all Grade X students enrolled at SMK Negeri 1 Sei Suka during the first semester of the 2026/2027 academic year. Purposive sampling is specified, with selected classes required to have comparable average mathematics examination scores and homogeneous initial variance. Under this design, the Grade X Office Management and Business Services (MPL) class is assigned to the experimental condition (n=36), while the Grade X Accounting and Institutional Finance (AKL) class is assigned to the control condition (n=36), yielding a total of 72 student records in the case-study dataset.

B. Instructional Materials

The worksheet package consists of six units, with a modelled instructional dosage of six meetings of two lesson periods each. Task contexts include administration, document services, transactions, workload, administrative packages, and problem posing. All tasks are designed to be self-contained so that students do not need unstated knowledge specific to MPL or AKL in order to solve them. The level of scaffolding decreases from Unit 1 (high support) to Unit 6 (minimal support), while each unit incorporates an explicit modelling cycle and a deliberate error/model-validation task. Unit 6 serves as a transfer task and includes problem posing with peer validation.

Table 1. Conceptual architecture of the CTL- and PBL-based student worksheets

Component	Primary Function	Operationalization in the Worksheets
CTL / REACT	Connects mathematics with context and transfer	Relating, Experiencing, Applying, Cooperating, Transferring
PBL	Organizes the social-pedagogical problem-based process	Problem orientation, learning organization, inquiry, presentation, evaluation
Polya framework	Organizes the cognitive problem-solving process	Understanding, planning, solving, looking back
Mathematical modelling cycle	Links real situations and mathematical models	Situation -> assumptions -> formulate -> solve -> interpret -> validate/revise
Metacognitive scaffolding	Regulates support and learner independence	Prompts, self-explanation, peer verification, progressive fading

C. Instrument and Scoring

The outcome instrument consists of five open-ended essay items. Each item is scored on four Polya indicators: Understanding the Problem (M), Planning the Solution (R), Solving the Model (S), and Looking Back/Checking (C), each rated on a 0-4 scale. The maximum score per item is 16, and the maximum raw test score is 80. Scores on a 0-100 scale are calculated as (raw score/80) x 100. The deliberate error task does not introduce a new measurement dimension; rather, it enriches the evidence for the Looking Back/Checking indicator and, where relevant, the Planning the Solution indicator.

The instrument audit is intended to ensure that all items contain internally consistent systems of linear equations, feasible solutions, and contexts that do not require hidden vocational knowledge. A double-scoring protocol is prespecified for empirical implementation using two raters, with a two-way random-effects, absolute-agreement, single-measure intraclass correlation coefficient [ICC(2,1)] as the primary index of inter-rater reliability.

D. Pilot Study, Response Process, and Content Validity

The pilot case study uses 30 synthetic respondents, comprising 15 MPL profiles and 15 AKL profiles, with 600 primary scores. Internal consistency is estimated using Cronbach's alpha [5] and corrected item-total correlations (CITC). In addition, 10 scripts are used to demonstrate double-scoring calibration (200 paired ratings), and six simulated cognitive-check profiles are used to illustrate response-process evidence. Content validity is demonstrated through four simulated validator profiles representing Mathematics Education/Mathematical Modelling, Instructional Design, an MPL practitioner, and an AKL practitioner. Aiken's V [1] is used as a summary index of agreement regarding content relevance and appropriateness.

E. Implementation Fidelity

Implementation fidelity is not reduced to a single percentage. Four modelled dimensions are used: adherence, dosage/time, quality of delivery, and participant responsiveness. Some sessions are modelled with co-observation to demonstrate calibration and agreement. This distinction follows fidelity literature showing that adherence, dosage, and quality are frequently treated as separate dimensions and should preferably be planned from the outset of a study [14].

F. Statistical Analysis

The hierarchy of analyses is prespecified before inspection of the final dataset. The primary demonstration analysis is an ANCOVA of posttest scores with pretest score as a covariate. The model is specified as POST_i = beta0 + beta1(Group_i) + beta2(PRE_i - PRE_mean) + epsilon_i. Because pretest scores are centered at the grand mean, beta1 represents the adjusted posttest difference at the average level of initial ability. Raw change (posttest minus pretest) is treated as a secondary analysis, while individual N-Gain is treated as a supporting analysis. Welch's test, the Mann-Whitney test, and HC3 robust covariance are used as sensitivity/robustness checks. Analyses of the four Polya indicators are exploratory, and p-values, when reported, are adjusted using the Holm procedure.

Diagnostic checks include homogeneity of regression slopes, linearity, residual normality, between-group variance, and influential cases. Baseline p-values are not used as evidence of group equivalence. Effect sizes are reported together with confidence intervals and with explicit identification of the outcome on which the calculation is based [8].

3. RESULTS AND DISCUSSION

A. Descriptive and Primary Results

Table 2. Descriptive statistics of mathematical problem-solving scores (simulated dataset)

Group	N	Pretest Mean ± SD	Posttest Mean ± SD
Experimental / MPL	36	45.451 ± 13.117	66.701 ± 14.218
Control / AKL	36	45.521 ± 10.528	56.354 ± 14.256

The pretest means of the two conditions are numerically very similar. This similarity is reported only as a descriptive characteristic of the measured dataset; a p-value greater than .05 is not used to claim that the two classes/programs are equivalent. Before adjustment, the descriptive posttest difference is 10.347 points.

Table 3. Results of the primary demonstration ANCOVA

Parameter	Value
Adjusted mean, MPL	66.733
Adjusted mean, AKL	56.323
Adjusted difference	10.410 points
95% CI	[5.947, 14.872]
F test	F(1,69)=21.656
p-value	<.001
Partial eta-squared	.239
Adjusted d (residual SD)	1.097

In the simulated dataset, the adjusted posttest mean for the MPL condition is 10.410 points higher than that of the AKL condition at the mean pretest level. The partial eta-squared and adjusted d indicate a large demonstrative difference within the case-study analysis pipeline. However, this value must not be interpreted as a causal treatment effect because the treatment is assigned to only one class/program; consequently, the group effect remains confounded with the class/program effect.

B. Diagnostics and Robustness

No indication of a group x pretest interaction is observed in the simulated dataset (p=.741), suggesting no evident violation of the homogeneity-of-regression-slopes assumption. The Shapiro-Wilk test of residuals yields p=.032, indicating a mild departure from normality, whereas Levene's test on residuals yields p=.502, indicating no evident between-group difference in residual variance. Three cases exceed the heuristic threshold Cook's D > 4/N, with a maximum Cook's D of .140; no case is removed solely on the basis of this heuristic.

Table 4. Secondary, supporting, and robustness analyses

Analysis	Summary of Results	Interpretation
HC3 robust ANCOVA	Adjusted diff=10.410; SE=2.273; 95% CI [5.875, 14.945]; p<.001	Consistent with primary OLS ANCOVA
Raw change (secondary)	MPL=21.250; AKL=10.833; diff=10.417; 95% CI [5.948, 14.885]; Welch t(67.85)=4.652; p<.001; g=1.085	Supports the primary pattern
N-Gain (supporting)	MPL=.4019; AKL=.2036; diff=.1983; 95% CI [.1092, .2874]; Welch t=4.441; p<.001	Supporting metric only
Mann-Whitney N-Gain	U=1006.5; p<.001	Sensitivity check because control-group N-Gain is non-normal

C. Changes in Polya Indicators

Table 5. Changes in Polya indicators (exploratory)

Indicator	Delta MPL	Delta AKL	Difference	Holm p
Understanding the problem	3.639 ± 2.140	2.333 ± 2.473	1.306	.019
Planning the solution	4.139 ± 2.086	2.528 ± 2.513	1.611	.008
Solving the model	5.056 ± 2.229	1.417 ± 2.209	3.639	<.001
Looking back/checking	4.167 ± 2.197	2.389 ± 2.142	1.778	.003

The largest numerical improvement in the MPL condition is observed for the Solving the Model indicator. Because the indicator-level analyses are exploratory, this pattern is used to support interpretation of the instructional design rather than as a primary confirmatory finding.

D. Measurement and Implementation Evidence

Table 6. Summary of measurement and implementation evidence (simulated demonstration)

Component	Main Result	Interpretive Boundary
Pilot internal consistency	Cronbach's alpha=.775; standardized alpha=.771; mean inter-item r=.403; CITC=.365-.739	Demonstration of internal consistency; not construct validity
B2 monitoring	CITC=.365; alpha-if-deleted=.786	Retain and monitor because of MPL content coverage
IRR calibration	Exact agreement=83.0%; within 1 point=100%; ICC(A,1)=.976	Calibration demonstration; not final empirical IRR
Instrument validity	Mean=4.58; mean Aiken's V=.894	Four simulated validator profiles
Worksheet validity	Mean=4.56; mean Aiken's V=.891	Four simulated validator profiles
Multidimensional fidelity	Adherence=91.67%; dosage=98.52%; quality=90.28%; responsiveness=85.00%; composite=91.37%	Simulated implementation evidence

E. Interpretation and Conceptual Implications

The analysis pipeline shows an adjusted posttest difference of 10.410 points in the simulated dataset, and this pattern remains consistent across HC3 sensitivity analysis, raw change, N-Gain, and the nonparametric check. This coherence indicates that the measurement design and analytical hierarchy operate internally without producing the kinds of statistical contradictions identified in earlier versions of the study. Nevertheless, statistical robustness cannot resolve the structural confounding created by having one class/program in each condition. Therefore, the findings demonstrate how effectiveness could be reported quantitatively if real empirical evidence were available; they do not provide evidence that the worksheets causally improve the mathematical problem-solving ability of actual students at SMK Negeri 1 Sei Suka.

The Solving the Model indicator shows the largest exploratory change. This pattern is conceptually consistent with the design because, across the six units, students are repeatedly required to identify constraints, formulate systems of linear equations, select strategies, execute solutions, and validate the resulting models. The appropriate interpretation, however, is that the pattern may be consistent with vocational mathematical modelling scaffolding as a hypothesized mechanism. The case-study data cannot establish that progressive fading or deliberate error tasks caused changes in any particular indicator.

The contribution of the design does not lie in presenting an 'integration' of CTL, PBL, REACT, and Polya as a new theoretical framework. Rather, its contribution lies in an architecture that assigns a distinct function to each component and links these functions through a modelling cycle. CTL/REACT ensures contextualization and transfer; PBL provides a structure for inquiry; Polya provides a cognitive language for reasoning and scoring; the modelling cycle requires bidirectional translation between context and mathematical model; and metacognitive scaffolding regulates when prompts are provided and when they are withdrawn. This design responds to the gap identified in the systematic review of PBL in vocational mathematics, which noted the limited availability of workplace-contextualized learning materials and the continuing challenge of designing authentic problems [11].

Progressive fading is also supported theoretically by research on metacognition and scaffolding. Hidayat et al. [6] reported positive effects of metacognitive instruction on mathematics learning, whereas Thomann and Deutscher [13] showed that prompts can improve achievement but that their effectiveness

depends on prompt type and implementation. Accordingly, the worksheets do not maintain detailed prompts throughout all six units; instead, support is gradually reduced until Unit 6, which requires problem posing, peer validation, and transfer. Problem posing itself is supported by meta-analytic evidence as a competence that can be enhanced through structured interventions [15].

F. Context Fairness, Validity, Fidelity, and Limitations

The shift in vocational specialization from technical contexts to MPL and AKL requires substantive reconstruction rather than simple renaming. The assessment items and worksheets are designed to be self-contained so that administrative, document-service, and transaction contexts function as sources of meaning rather than as sources of advantage based on domain-specific knowledge. Item B2 represents an MPL-related context, B3 approximates an AKL-related context, and the remaining items are cross-program in nature. The decision to retain B2 despite its lower CITC illustrates that instrument decisions should not be driven solely by maximizing alpha; construct coverage and fairness must also be considered.

Measurement evidence is separated into multiple sources. Cronbach's alpha and CITC describe internal consistency; Aiken's V summarizes agreement regarding content relevance; the calibration ICC represents the workflow for inter-rater reliability; and response-process checks examine whether terms and contexts are understood as intended. This separation avoids using a single statistic to represent the entire concept of validity. In an empirical implementation, all four sources of evidence would still need to be collected from actual students, raters, and validators.

Implementation fidelity is likewise treated as multidimensional. The literature shows variation in definitions of fidelity and commonly distinguishes adherence, dosage, and quality as separate dimensions [14]. Accordingly, the composite value of 91.37% is not interpreted in isolation. For example, the lower participant responsiveness relative to dosage illustrates that 'time completed' is not equivalent to student engagement. In an empirical study, such fidelity dimensions would be more informative for explaining implementation conditions than a single global score.

The principal limitations of the case study were as follows:

(1) All main data for 72 students, the 30-respondent pilot, validator ratings, rater calibration, response-process profiles, and fidelity evidence are simulated/reconstructed rather than obtained from actual students, validators, raters, or observers.

(2) Only one class/program is represented in each condition (MPL as experimental and AKL as control). Because the treatment is assigned at the class level, the treatment effect cannot be separated from the class/program effect; ANCOVA does not remove this confounding.

(3) N-Gain is sensitive to pretest scores and measurement error and is therefore used only as a supporting metric rather than as the primary outcome.

(4) The sample size and one-class-per-condition design do not support causal generalization. Future empirical research should include multiple classes per condition and, where feasible, use cluster randomization or a matched multilevel design.

4. CONCLUSION

Based on this quantitative methodological/academic case study, the use of CTL- and PBL-based student worksheets designed through vocational mathematical modelling scaffolding produces a coherent measurement and analysis pipeline. In the simulated dataset, the adjusted posttest difference between the MPL and AKL conditions is 10.410 points (95% CI [5.947, 14.872]), $F(1,69) = 21.656$, $p < .001$, partial eta-squared=.239, with an adjusted d of 1.097. Raw change, N-Gain, and robust sensitivity analyses show numerically consistent patterns, while the largest exploratory improvement is observed for the Solving the Model indicator.

The present study may serve as a prospective blueprint for empirical research. Before any claim of empirical effectiveness is made, future research should collect actual pretest-posttest data using the same finalized instrument, document the purposive-sampling process, conduct a real pilot study, obtain expert judgment from actual validators, perform double scoring on a subset of student responses, observe actual implementation fidelity across the four dimensions, and replicate the intervention across multiple classes. ANCOVA or mixed/multilevel models may then be selected in accordance with the resulting data structure. A delayed transfer task and analysis of students' written work are also recommended to determine whether modelling competence transfers to new contexts and to explain the process of change rather than focusing solely on outcome scores.

REFERENCES

[1] Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. Educational and Psychological Measurement, 45(1), 131-142. https://doi.org/10.1177/0013164485451012

[2] Barrows, H. S. (1986). A taxonomy of problem-based learning methods. Medical Education, 20(6), 481-486.
[3] Campbell, D. T., & Stanley, J. C. (1963). Experimental and Quasi-Experimental Designs for Research. Houghton Mifflin.
[4] Crawford, M. L. (2001). Teaching Contextually: Research, Rationale, and Techniques for Improving Student Motivation and Achievement in Mathematics and Science. CCI Publishing.
[5] Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297-334. https://doi.org/10.1007/BF02310555
[6] Hidayat, R., Mohd Saad, M. R., & Wewe, M. (2025). A meta-analysis of the effect of metacognitive instruction on mathematics achievement. Cogent Education, 12(1), 2517510. https://doi.org/10.1080/2331186X.2025.2517510
[7] Johnson, E. B. (2002). Contextual Teaching and Learning: What It Is and Why It Is Here to Stay. Corwin Press.
[8] Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Frontiers in Psychology, 4, 863. https://doi.org/10.3389/fpsyg.2013.00863
[9] OECD. (2023). PISA 2022 Assessment and Analytical Framework. OECD Publishing.
[10] Polya, G. (1973). How to Solve It: A New Aspect of Mathematical Method (2nd ed.). Princeton University Press.
[11] Rendrayana, K., Suparta, I. N., & Suharta, I. G. P. (2026). Problem-Based Learning dalam Matematika Sekolah Menengah Kejuruan (SMK) Berbasis Masalah Kontekstual: Systematic Literature Review. Media Pendidikan Matematika, 14(1), 296-310. https://doi.org/10.33394/mpm.v14i1.20568
[12] Salsabila, S., & Asih, E. C. M. (2024). The Effect of Problem-Based Learning Models on Students' Mathematical Problem-Solving Ability: A Meta-Analysis. Jurnal Pendidikan MIPA, 25(2), 864-877. https://doi.org/10.23960/jpmipa/v25i2.pp864-877
[13] Thomann, H., & Deutscher, V. (2025). Scaffolding through prompts in digital learning: A systematic review and meta-analysis of effectiveness on learning achievement. Educational Research Review, 47, 100686. https://doi.org/10.1016/j.edurev.2025.100686
[14] van Dijk, W., Lane, H. B., & Gage, N. A. (2023). How Do Intervention Studies Measure the Relation between Implementation Fidelity and Students' Reading Outcomes? A Systematic Review. The Elementary School Journal, 124(1), 56-84. https://doi.org/10.1086/725672
[15] Zhang, L., Stylianides, G. J., & Stylianides, A. J. (2024). Enhancing mathematical problem posing competence: A meta-analysis of intervention studies. International Journal of STEM Education, 11, 48. https://doi.org/10.1186/s40594-024-00507-1