HYDROGEN
JURNAL KEPENDIDIKAN KIMIA
https://e-journal.undikma.ac.id/index.php/hydrogen/index

June 2026. 14(3)
e-ISSN: 2656-3061
p-ISSN: 2338-6487
pp. 579-592

Assessment-for-Learning-Based Student Activity Sheets on Reaction Rate to Improve Students' Science Process Skills

Moh. Threewahyu Saifulloh & Rudiana Agustini*
Chemistry Department, Faculty of Mathematics and Natural Science, Universitas Negeri Surabaya, Jl. Ketintang, Ketintang, Kec. Gayungan, Surabaya, Indonesia
*Corresponding Author e-mail: rudianaagustini@unesa.ac.id

Article History
Received: 30-05-2026
Revised: 18-06-2026
Published: 30-06-2026

Keywords: Student Activity Sheets; Assessment for Learning; Guided Inquiry; Science Process Skills; Reaction Rate.

Abstract
Science process skills remain underdeveloped in many chemistry classrooms due to limited use of structured, feedback-oriented learning materials. This study aimed to develop Student Activity Sheets (*Lembar Aktivitas Murid*/LAM) based on Assessment for Learning (AfL) principles to improve students' science process skills in reaction rate material. This study used a Research and Development design based on the 4D model, limited to the Define, Design, and Develop stages. The product was validated by two chemistry education lecturers and one chemistry teacher. Practicality was evaluated through student response questionnaires and activity observation sheets, while effectiveness was examined using a one-group pretest-posttest design involving 34 grade XI students at SMAN 1 Menganti. The results showed that the LAM was valid, with a mode score of 4 across content, construct, and language validity aspects. The product was also very practical, with student response and observation scores of 98.53% and 89.25%, respectively. The Wilcoxon signed-rank test confirmed a statistically significant improvement between pretest and posttest scores ($Z = -5.093$, $p < .001$), with a large effect size ($r = 0.87$). N-gain analysis revealed an overall mean gain of 0.73 (high category), with four indicators in the high category — collecting data (0.83), identifying variables (0.81), analyzing data (0.78), and formulating problems (0.74) — and two in the medium category — drawing conclusions (0.63) and constructing hypotheses (0.61). These findings indicate that the AfL-based LAM is feasible and effectively supports the development of students' science process skills in reaction rate learning.

How to Cite: Saifulloh, M. T., & Agustini, R. (2026). Assessment-for-Learning-Based Student Activity Sheets on Reaction Rate to Improve Students' Science Process Skills. Hydrogen: Jurnal Kependidikan Kimia, 14(3), 579-592. https://doi.org/10.33394/hjkk.v14i3.20933

https://doi.org/10.33394/hjkk.v14i3.20933

INTRODUCTION

Education serves as a structured process aimed at fostering students' holistic development, encompassing intellectual, attitudinal, and practical competencies essential for navigating everyday life (Maskur, 2023). In Indonesia, the Merdeka Curriculum (Kepmendikbudristek, 2022) has shifted educational priorities toward learner-centered approaches, contextual learning experiences, and ongoing assessment of the learning journey rather than summative outcomes alone (Maulani et al., 2024). Within this framework, chemistry stands as a particularly challenging subject due to the abstract nature of its concepts, reliance on mathematical reasoning, and instructional approaches that remain insufficiently engaging (Pratomo et al., 2025). Teacher-directed methods have been associated with diminished student involvement and reduced depth of learning (Hakim et al., 2024). Pre-research data collected through questionnaires at SMAN 1 Menganti confirmed this condition, with 82.8% of students identifying chemistry as difficult, citing unclear material delivery, limited connection to everyday phenomena, and inadequate learning media.

Among chemistry topics in grade XI, reaction rate is consistently identified as conceptually demanding, with students frequently encountering significant barriers to understanding its underlying principles (Azmi et al., 2021).

Pre-research questionnaire results at SMAN 1 Menganti showed that 74.3% of students considered reaction rate difficult, while Arviani (2011) found that 70.8% of students held misconceptions about its core concepts. Reaction rate material inherently requires active student engagement in scientific inquiry, making it an appropriate context for developing science process skills (Basit et al., 2023).

Science process skills constitute a foundational set of competencies in chemistry education, empowering students to engage meaningfully in scientific inquiry and cultivate systematic scientific reasoning (Gürses et al., 2015). These skills encompass observation, classification, hypothesis formulation, experimental design, data analysis, and drawing conclusions (Tawil & Liliasari, 2014). However, pre-research diagnostic tests at SMAN 1 Menganti revealed that several science process skill indicators remained low in reaction rate material, including problem formulation (55.3%), hypothesis construction (43.9%), experiment planning (54.3%), and drawing conclusions (47.0%). These results indicate that students are not yet accustomed to engaging systematically in scientific processes during chemistry learning.

Student Activity Sheets (Lembar Aktivitas Murid/LAM) represent structured instructional materials designed to guide students in understanding concepts through direct involvement in the learning process (Prastowo, 2015). LAM has the potential to improve science process skills when designed systematically and oriented toward scientific activities (Aldiyah, 2021). Despite 80% of students at SMAN 1 Menganti already using LAM—based on pre-research questionnaire data—learning remained dominated by lecture methods (77.1%), while 80% of students preferred discussion-based approaches. This indicates that existing LAM had not been optimized as a tool for active learning and science process skill development.

To address this limitation, integration with Assessment for Learning (AfL) is proposed. AfL is a formative evaluation framework in which assessment is embedded within the instructional process, functioning as a continuous mechanism to enhance learning quality through systematic feedback (Assessment Reform Group, 2002; Budiyono, 2011). Previous studies support this approach: Fitrianna and Anita (2017) reported that AfL-based LAM enhanced student interaction, while Utami and Agustini (2024) confirmed that it promoted active engagement and contributed to significant improvement in science process skills. Nurfidayanti and Yonata (2022) further demonstrated that AfL integration supported the systematic development of process skill indicators, including problem formulation and hypothesis construction. Black and Wiliam (2009) emphasized that formative feedback enables students to identify their learning position and independently plan corrective steps.

However, existing studies have not specifically examined the comprehensive integration of all five AfL strategic steps within LAM designed to systematically develop seven science process skill indicators in reaction rate material. The five AfL steps—establishing clear learning objectives and success criteria, designing purposeful learning tasks that generate evidence of comprehension, delivering forward-oriented feedback, fostering peer learning, and cultivating student autonomy—have not been holistically mapped onto the seven process skill indicators (formulating problems, constructing hypotheses, identifying variables, planning experiments, collecting data, analyzing data, and drawing conclusions) within a guided inquiry model at the senior secondary level. Pre-research interview and questionnaire data at SMAN 1 Menganti further support this gap: only 60% of students had prior exposure to LAM with structured feedback, while 85.7% perceived AfL as a promising approach for improving science process skills.

Grounded in this rationale, this study addresses the following research questions: (1) How valid is the AfL-based LAM developed for reaction rate material? (2) How practical is the LAM based on student responses and activity observations? (3) How effective is the LAM in improving students' science process skills? Accordingly, this study aims to develop and evaluate the feasibility of Student Activity Sheets incorporating AfL principles in reaction rate material, assessed across three dimensions—validity, practicality, and effectiveness—following the evaluative framework proposed by Plomp and Nieveen (2013), with the overarching goal of enhancing the science process skills of grade XI students at SMAN 1 Menganti.

METHOD

Research Design

The present study adopted a Research and Development (R&D) methodology, utilizing the four-stage development model introduced by Thiagarajan et al. (1974), which encompasses the Define, Design, Develop, and Disseminate phases. However, the scope of this study was confined to the initial three stages—Define, Design, and Develop—forming a modified 3D model. The Disseminate stage was excluded due to practical constraints related to time and resource limitations within the scope of this study, as the primary objective was to evaluate the validity, practicality, and effectiveness of the developed product rather than to distribute it on a broader scale. Future research may extend this work by implementing the Disseminate stage across multiple school settings.

To evaluate effectiveness, a One Group Pretest-Posttest Design was employed, in which a single group of students was assessed before and after LAM implementation without a control group. This design was selected given the exploratory nature of the study and the availability of only one intact class at the research site. It is acknowledged that this design carries an inherent limitation: the absence of a control group means that observed improvements in students' science process skills scores after implementation cannot be attributed exclusively to the LAM, as other variables may have contributed to the changes. Accordingly, findings related to effectiveness are interpreted cautiously—that students' science process skills scores improved following LAM implementation—rather than as definitive causal claims.

The product developed in this study is Student Activity Sheets (Lembar Aktivitas Murid/LAM) based on Assessment for Learning (AfL) principles, focused on reaction rate material—specifically the sub-topic of factors affecting reaction rate—designed to improve the science process skills of grade XI senior secondary school students. The research flow is presented in Figure 1.

Research Subject

The primary participants of this study comprised 34 students enrolled in class XI-1 at SMAN 1 Menganti, consisting of 14 male and 20 female students with an age range of 16 to 17 years. This class was selected purposively in coordination with the collaborating chemistry teacher, as class XI-1 represented an intact class with a sufficient number of students for a limited trial and reflected typical learning conditions at the school, including low achievement on several science process skill indicators in reaction rate material based on pre-research results.

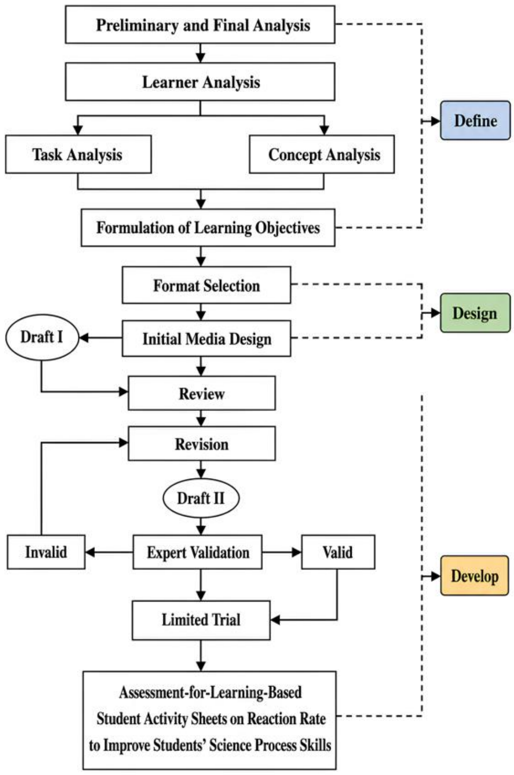

Figure 1. Research Flow of 4D Development Model (Modified into 3D)

The validity assessment involved three validators: two chemistry education lecturers with expertise in instructional material development and chemical assessment, and one practising chemistry teacher with a minimum of five years of experience teaching reaction rate material at the senior secondary level. These validators were selected based on their competence in evaluating both the content accuracy and construct alignment of the developed LAM with AfL principles and science process skill indicators.

The practicality assessment was carried out with the involvement of six observers who completed structured activity observation sheets during LAM implementation. Prior to data collection, all observers received the same observation guidelines and a briefing session to ensure consistency and uniformity in how student activities were recorded and interpreted across all observation sheets.

Development Procedure

Define Stage

The Define stage was intended to determine the foundational prerequisites for LAM development through five analytical procedures. First, front-end analysis was conducted to diagnose core challenges in chemistry learning at SMAN 1 Menganti through direct classroom observation and pre-research questionnaires administered to 35 students of class XII-1. Second, learner analysis examined student profiles encompassing academic capability, cognitive maturity, motivational orientation, and baseline science process skill proficiency. Third, task analysis delineated the learning tasks necessary for cultivating science process skills in alignment with the AfL strategic framework. Fourth, concept analysis mapped the essential conceptual content of reaction rate material for incorporation into the LAM. Fifth, instructional objective specification was derived from the consolidated findings of the task and concept analyses to establish the targeted learning competencies underpinning LAM development.

Design Stage

The Design stage aimed to produce an initial LAM design through two steps. Format selection determined the appropriate LAM structure, layout, and presentation format aligned with AfL principles and science process skill indicators. Initial design then produced a draft LAM incorporating all five AfL strategic steps integrated with seven science process skill indicators, operationalized through a guided inquiry learning model. The mapping between AfL strategic steps, science process skill indicators, and LAM activities is presented in Table 1.

Table 1. Mapping of AfL Strategic Steps, Science Process Skill Indicators, and LAM Activities

AfL Strategic Step	Science Process Skill Indicator	LAM Activity
Clarifying learning objectives and success criteria	Formulating problems	Students read contextual phenomena and formulate investigable research questions
Engineering effective classroom discussions and learning tasks	Constructing hypotheses; Identifying variables	Students propose hypotheses and identify independent, dependent, and controlled variables through guided prompts
Providing feedback	Planning experiments; Collecting data	Students design experimental procedures and collect data; written feedback is embedded in structured reflection boxes
Activating students as learning resources for peers	Analyzing data	Students discuss and cross-check data interpretation results with peers through guided peer-review tasks
Activating students as owners of their own learning	Drawing conclusions	Students independently formulate conclusions and evaluate their own learning progress through self-assessment checklists

In total, four LAM units were developed, each corresponding to one meeting and one factor affecting reaction rate. LAM Unit 1 addresses the effect of concentration on reaction rate, LAM Unit 2 addresses surface area, LAM Unit 3 addresses temperature, and LAM Unit 4 addresses catalyst. Each unit is structured consistently, embedding all five AfL strategic steps and targeting the same seven science process skill indicators across different experimental contexts, thereby allowing progressive development of students' science process skills throughout the learning sequence.

Develop Stage

The Develop stage consisted of three sequential steps. Expert validation was first conducted by three validators to assess the content and construct validity of the developed LAM using Likert scale validation forms. Based on validator feedback, revisions were made before proceeding to the limited trial. The limited trial was subsequently implemented with 34 students of class XI-1 at SMAN 1 Menganti across four meetings to assess practicality and collect pretest-posttest data for effectiveness measurement, following the One Group Pretest-Posttest Design ($O_1 \times O_2$), where O_1 represents the pretest administered prior to LAM implementation and O_2 represents the posttest administered after LAM implementation. Final revisions were then carried out based on limited trial findings.

Data Collection Instruments

Several instruments were employed in this study to evaluate the validity, practicality, and effectiveness of the developed LAM. An overview of all instruments is presented in Table 2.

Table 2. Instrument Specification Summary

Instrument	Purpose	Number of Items	Scale/Format	Respondents	Data Analysis
Likert scale validation form (content & construct validity)	Assess content accuracy, construct alignment with AfL and science process skill indicators, and language clarity	17	Likert scale (1–5)	2 chemistry lecturers, 1 chemistry teacher	Quantitative descriptive (mode/median)
Student response questionnaire	Measure perceived practicality of LAM	22	Guttman scale (Yes = 1 / No = 0 for positive items; reversed for negative items)	34 students	Quantitative descriptive (percentage)
Activity observation sheet	Measure observable student activity during LAM implementation	11	Rating scale (percentage of activities carried out)	6 observers	Quantitative descriptive (percentage)
Science process skills test (pretest-posttest)	Measure effectiveness of LAM in improving science process skills	7	Open-ended essay (7 items, 1 per indicator)	34 students	Wilcoxon signed-rank test and N-gain

Validity Instrument

Content validity, construct validity, and language quality were assessed using a Likert scale validation form employing a five-point scale (1 = very invalid, 2 = invalid, 3 = moderately valid, 4 = valid, 5 = very valid), adapted from Riduwan (2015). Validity was determined based on the mode score across all items, with a mode of ≥ 4 indicating the aspect is valid. In the absence of a mode—where all validator scores differ—the median was used as an alternative measure. If the mode or median score was < 4, revision and re-validation were conducted until met (Lutfi, 2021).

Practicality Instruments

Practicality was assessed through two instruments. The student response questionnaire used a Guttman scale with dichotomous response options (yes/no), comprising positive and negative statement items. For positive items, "yes" was scored 1 and "no" was scored 0; scoring was reversed for negative items. The activity observation sheet recorded the percentage of student learning activities carried out during LAM implementation across four meetings. Both instruments were analyzed by calculating the percentage of total scores against the maximum possible score, with $\geq 61\%$ categorized as practical based on Riduwan's (2015) classification.

Effectiveness Instrument

Effectiveness was measured using a science process skills test administered as pretest and posttest, consisting of seven open-ended essay items, each corresponding to one of the seven science process skill indicators. Each item was scored using a four-point rubric with specific answer criteria as presented in Table 3.

Table 3. Scoring Rubric for Science Process Skills Test

No.	Indicator	Answer Criteria	Score Description
1	Formulating Problems	(1) Written as a question; (2) corresponds to the phenomenon; (3) contains two related variables (independent and response variables)	4 = meets all three criteria accurately; 3 = meets two criteria; 2 = meets one criterion; 1 = meets no criterion

No.	Indicator	Answer Criteria	Score Description
2	Constructing Hypotheses	(1) Written as an if-then-therefore statement; (2) corresponds to the phenomenon; (3) states the relationship between independent and response variables	4 = meets all three criteria logically and testably; 3 = meets two criteria; 2 = meets one criterion; 1 = meets no criterion
3	Identifying Variables	(1) Correctly identifies independent variable; (2) correctly identifies control variable; (3) correctly identifies response variable	4 = identifies all three variable types correctly and completely; 3 = identifies two types correctly; 2 = identifies one type correctly; 1 = unable to identify variables correctly
4	Planning Experiments	(1) Determines appropriate tools and materials; (2) arranges work steps systematically; (3) steps align with the experimental objective	4 = complete, systematic, and aligned with objective; 3 = sufficiently systematic but incomplete; 2 = insufficiently systematic and misaligned; 1 = unable to plan the experiment correctly
5	Collecting Data	(1) Completes observation table fully; (2) data corresponds to experimental results; (3) data is recorded systematically	4 = meets all three criteria completely; 3 = meets two criteria; 2 = meets one criterion; 1 = data is inaccurate or incomplete
6	Analyzing Data	(1) Answers all analysis questions correctly; (2) connects data to relevant reaction rate concepts; (3) provides logical explanations aligned with the concept	4 = complete, logical, and conceptually accurate; 3 = sufficiently accurate but lacks depth; 2 = inaccurate and incomplete; 1 = unable to analyze data correctly
7	Drawing Conclusions	(1) Written as a declarative statement; (2) aligned with the learning objective; (3) consistent with the problem formulation and data	4 = meets all three criteria accurately; 3 = meets two criteria; 2 = meets one criterion; 1 = meets no criterion

Score conversion: Science Process Skills Score = (Total score obtained / Maximum total score) × 100

The science process skills test was validated by the same three expert validators prior to use. Pretest and posttest scores were analyzed using two complementary techniques. First, the Wilcoxon Signed-Rank Test was conducted using IBM SPSS Statistics (α = 0.05) to determine whether a statistically significant difference existed between pretest and posttest scores overall. LAM was considered effective if $p < 0.05$. Second, the normalized gain (N-gain) score was calculated for each of the seven science process skill indicators individually to measure the magnitude of improvement per indicator, using the formula proposed by Hake (1999):

$$N - gain = \frac{S_{post} - S_{pre}}{S_{max} - S_{pre}}$$

where S_{post} is the posttest score, S_{pre} is the pretest score, and S_{max} is the maximum possible score. The N-gain score for each indicator was interpreted based on the criteria presented in Table 4 (Hake, 1999).

Table 4. N-Gain Score Interpretation Criteria

N-Gain Score	Category
$g > 0.70$	High
$0.30 \le g \le 0.70$	Medium
$g < 0.30$	Low

N-gain analysis was conducted separately for each science process skill indicator—formulating problems, constructing hypotheses, identifying variables, planning experiments, collecting data, analyzing data, and drawing conclusions—to provide a detailed profile of improvement across all indicators. This analysis complements the Wilcoxon test by not only confirming the statistical significance of overall improvement but also revealing which indicators demonstrated the greatest and least gains following LAM implementation.

Data Analysis Technique

Validity Analysis

Validity data were analyzed using quantitative descriptive analysis covering three aspects: content, construct, and language validity, assessed by two chemistry education lecturers and one chemistry teacher using a Likert scale as presented in Table 5 (Adapted from Riduwan, 2015)

Table 5. Likert Scale Criteria for Validity Assessment

Score	Category
1	Invalid
2	Less Valid
3	Fairly Valid
4	Valid
5	Very Valid

Analysis was performed by determining the mode score for each aspect. If the mode score was ≥ 4, the aspect was declared valid; if < 4, revision was required. When no mode existed — that is, when all validator scores differed—the median was used as an alternative measure of central tendency, given the ordinal nature of Likert scale data (Riduwan, 2015). The validity decision criteria are presented in Table 6 (Lutfi, 2021).

Table 6. Validity Decision Criteria

Mode / Median Score	Decision
≥ 4	Valid
< 4	Invalid (revision required)

Practicality Analysis

Practicality was assessed through two instruments: student response questionnaires and activity observation sheets. Student responses were measured using a Guttman scale with dichotomous answer choices (yes/no) as presented in Table 7. Positive statement items were those affirming ease of use, clarity of instructions, and active engagement during learning—for example, "The instructions in the LAM are easy to understand." Negative statement items were those indicating difficulty or disengagement—for example, "The activities in the LAM are confusing and hard to follow".

Table 7. Guttman Scale for Student Response

Statement	Response	Score
Positive	Yes	1
	No	0
Negative	Yes	0
	No	1

(Riduwan, 2015)

Both student response and activity observation data were analyzed by calculating the percentage score using the following formula:

$$\%Practically = \frac{\Sigma \text{ scores obtained}}{\Sigma \text{ maximum scores}} \times 100\%$$

The percentage results were interpreted based on the criteria presented in Table 8.

Table 8. Practicality Interpretation Criteria

Percentage Score (%)	Category
0 – 20	Very Impractical
21 – 40	Impractical
41 – 60	Fairly Impractical
61 – 80	Practical
81 – 100	Very Practical

(Riduwan, 2015)

Effectiveness Analysis

Prior to inferential analysis, a normality test was conducted on both pretest and posttest scores using the Shapiro-Wilk test, which is appropriate for sample sizes of n ≤ 50 (Razali & Wah, 2011). The normality test results are presented in Table 9.

Table 9. Tests of Normality for Pretest and Posttest Scores

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pretest	.185	34	.005	.881	34	.001
Posttest	.185	34	.004	.929	34	.029

a. Lilliefors Significance Correction

The Shapiro-Wilk test results showed that both pretest and posttest scores were not normally distributed. The pretest scores yielded a Shapiro-Wilk statistic of .881 (df = 34, p = .001), and the posttest scores yielded a Shapiro-Wilk statistic of .929 (df = 34, p = .029). Both p-values fell below the significance threshold of α = .05, indicating that the normality assumption was violated for both datasets. This finding was further corroborated by the Kolmogorov-Smirnov test, which also yielded significant results for both pretest (statistic = .185, df = 34, p = .005) and posttest (statistic = .185, df = 34, p = .004). The violation of the normality assumption in both pretest and posttest data necessitated the use of a non-parametric statistical test. Accordingly, the

Wilcoxon Signed-Rank Test was selected as the appropriate analytical procedure for comparing pretest and posttest scores (Sugiyono, 2019).

The Wilcoxon Signed-Rank Test was conducted using IBM SPSS Statistics with a predetermined significance threshold of $\alpha = .05$. A p-value below .05 was interpreted as evidence of a statistically significant difference between pretest and posttest scores, indicating that LAM implementation was associated with meaningful improvement in students' science process skills in reaction rate material. The effectiveness decision criteria are presented in Table 10 (Sugiyono, 2019).

Table 10. Effectiveness Decision Criteria Based on Wilcoxon Test

Significance Value (p-value)	Decision	Interpretation
$p < 0.05$	Reject H_0	Significant difference between pretest and posttest; LAM is effective
$p \geq 0.05$	Accept H_0	No significant difference between pretest and posttest; LAM is not effective

The Wilcoxon Signed-Rank Test results are presented in Table 1.

Table 11. Ranks and Test Statistics of Wilcoxon Signed-Rank Test

Ranks

		N	Mean Rank	Sum of Ranks
Posttest - Pretest	Negative Ranks	0[a]	.00	.00
	Positive Ranks	34[b]	17.50	595.00
	Ties	0[c]		
	Total	34		

a. Posttest < Pretest
b. Posttest > Pretest
c. Posttest = Pretest

Test Statistics[a]

	Posttest - Pretest
Z	-5.093[b]
Asymp. Sig. (2-tailed)	.000

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

As shown in Table 11, all 34 students (100%) showed positive ranks, indicating that posttest scores were higher than pretest scores for every participant, with no negative ranks or ties observed. The Test Statistics in Table 12 revealed a Z-value of −5.093 with an asymptotic significance of $p < .001$, which is well below the threshold of $\alpha = .05$. Therefore, H_0 was rejected, confirming a statistically significant difference between pretest and posttest scores. This result indicates that students' science process skills improved significantly following LAM implementation.

To quantify the strength of this improvement, effect size r was calculated using the formula proposed by Cohen (2013):

$$r = \frac{|Z|}{\sqrt{N}} = \frac{5.093}{\sqrt{34}} = \frac{5.093}{5.831} = 0.87$$

The effect size r = 0.87 falls in the large category ($r \geq 0.50$) based on the criteria presented in Table 12, indicating that the improvement in students' science process skills was not only statistically significant ($p < .001$) but also practically meaningful in magnitude (Cohen, 2013).

Table 12. Effect Size Interpretation Criteria

Effect Size (r)	Category
$r < 0.10$	Negligible
$0.10 \leq r < 0.30$	Small
$0.30 \leq r < 0.50$	Medium
$r \geq 0.50$	Large

RESULTS AND DISCUSSION

Validity of the LAM

Content and Construct Validity

The developed LAM underwent a validation process involving three validators — two chemistry education lecturers and one experienced chemistry teacher — covering three evaluative dimensions: content validity, construct validity, and language validity. The LAM consistently attained a mode score of 4 across all three dimensions, categorized as valid in accordance with the predetermined assessment criteria (Lutfi, 2021). The validity results are presented in Table 13.

The mode score of 4 across all aspects confirms that the LAM satisfies validity standards with respect to subject matter alignment, structural coherence, and linguistic appropriateness.

Table 123. Validity Assessment Results

No	Validity Aspect	Mode Score	Category
1	Content Validity	4	Valid
2	Construct Validity	4	Valid
3	Language Validity	4	Valid
	Overall	4	Valid

Content validity confirms that the reaction rate material — particularly factors affecting reaction rate — is scientifically accurate, contextually relevant, and aligned with Merdeka Curriculum competency standards. Construct validity confirms that the five AfL strategic steps are logically integrated with the seven science process skill indicators, consistent with Nurfidayanti and Yonata (2022), who established that embedding AfL steps within structured learning activities supports the advancement of science process skills. Language validity confirms that instructions, questions, and feedback components are written in language comprehensible and appropriate for grade XI students.

In addition to quantitative scores, validators provided qualitative suggestions that served as the basis for product revision prior to the limited trial. The validator suggestions and corres-ponding revisions are presented in Table 14.

Table 134. Summary of Validator Suggestions and Product Revisions

No	Validator Suggestion	Revised Part	Revision Result
1	Scientific terminology in several activity items was inconsistent with standard chemistry nomenclature	Scientific terminology throughout all four LAM units	Terminology was standardized to align with IUPAC conventions and Merdeka Curriculum glossary
2	Experimental procedure instructions were insufficiently detailed, potentially causing confusion during practical activities	Step-by-step procedure sections in LAM Units 1–4	Instructions were elaborated with specific quantities, sequential numbering, and safety notes
3	Feedback guidance prompts were too general and did not direct students toward specific self-correction	Written feedback boxes embedded after each activity section	Feedback prompts were rewritten to include targeted guiding questions aligned with each science process skill indicator
4	Learning objectives were not explicitly stated at the opening of each LAM unit	Learning objective section at the beginning of each unit	Learning objectives were reformulated using observable behavioral verbs and explicitly linked to success criteria
5	Some contextual phenomena presented at the opening of LAM units were not sufficiently connected to the experimental activities that followed	Contextual phenomenon section in LAM Units 2 and 3	Phenomena were revised to directly relate to the experimental variables and guided inquiry questions in each unit

These revisions ensured that the final LAM was not only theoretically valid but also practically refined for classroom use. The systematic development process following the Define, Design, and Develop stages of the 4D model (Thiagarajan et al., 1974) further supports the overall validity of the LAM, as the thorough front-end, learner, task, and concept analyses conducted during the Define stage grounded the LAM in the actual learning needs of the target students, consistent with Prastowo (2015).

Practicality of the LAM

Student Response Questionnaire Results

Practicality assessment based on student response questionnaires was completed by 34 students of class XI-1 at SMAN 1 Menganti following the limited trial. To minimize response bias, the questionnaire was administered anonymously after all learning activities had concluded, and students were explicitly informed that their responses would not affect their academic grades. The questionnaire employed a Guttman scale (yes/no) comprising both positive and negative statement items to reduce acquiescence bias. Examples of positive items included "The instructions in the LAM are easy to understand" and "The contextual phenomena presented in the LAM are relevant to everyday life," while examples of negative items included "The activities in the LAM are confusing and difficult to follow." Results showed an overall practicality percentage of 98.53%, categorized as very practical. The detailed results are presented in Table 15.

Table 145. Student Response Questionnaire Results

No	Assessed Aspect	Percentage (%)	Category
1	Clarity of learning objectives	97,06	Very Practical
2	Clarity of activity instructions	98,53	Very Practical
3	Relevance of contextual phenomena	100	Very Practical
4	Feedback usefulness	98,53	Very Practical
5	Encouragement of active participation	97,06	Very Practical
6	Overall LAM appearance and layout	100	Very Practical
	Overall	98,53	Very Practical

Despite the measures taken to reduce bias, it is acknowledged that the very high student response score of 98.53% may still reflect a degree of social desirability bias, as students may have tended toward positive responses when evaluating materials used in their own classroom. This limitation should be considered when interpreting the student response data, and the activity observation results — collected independently by trained observers — are therefore regarded as a more objective measure of practicality.

Activity Observation Sheet Results

Practicality assessment based on activity observation sheets was completed by six observers during the four-meeting limited trial. Results showed an overall practicality percentage of 89.25%, categorized as very practical. The detailed results are presented in Table 16.

Table 156. Activity Observation Sheet Results

No	Meeting	Percentage (%)	Category
1	Meeting 1	84	Very Practical
2	Meeting 2	87	Very Practical
3	Meeting 3	91	Very Practical
4	Meeting 4	95	Very Practical
	Overall	89,25	Very Practical

The gap between the student response score (98.53%) and the activity observation score (89.25%) is noteworthy and merits interpretation. Student response questionnaires capture students' subjective perceptions of the LAM, which tend to reflect overall satisfaction and perceived ease of use. In contrast, activity observation sheets record the actual execution of prescribed learning activities by trained external observers, making them less susceptible to subjective bias. The lower observation score relative to the student response score therefore suggests that while students perceived the LAM as highly clear and engaging, the actual implementation of certain activities — particularly those requiring more complex scientific reasoning such as experimental planning and data analysis — was not always fully completed within the allocated time. This discrepancy is consistent with findings by Fitrianna and Anita (2017), who noted that the effective implementation of AfL-based materials typically requires an initial adjustment period during which students and teachers gradually internalize the procedural framework.

The observation scores demonstrated a consistent upward trend across all four meetings, increasing progressively from Meeting 1 (84.00%) through Meeting 2 (87.00%), Meeting 3 (91.00%), to Meeting 4 (95.00%). This pattern strongly supports the interpretation that both students and the teacher became increasingly familiar and confident with the AfL-based learning procedures as the trial progressed, resulting in progressively more complete and systematic activity implementation across meetings. The steady improvement across all four meetings — with no decline at any point — indicates that the guided inquiry structure of the LAM successfully supported students in developing their science process skills in a scaffolded and progressive manner. The highest observation score at Meeting 4 (95.00%) suggests that by the final meeting, students had sufficiently internalized the AfL procedural framework to engage with the catalyst sub-topic activities with a high degree of independence and completeness.

Both instruments collectively confirm that the LAM can be implemented effectively in actual classroom conditions. The high practicality of the LAM is further attributable to its guided inquiry design, which scaffolded students progressively through scientific activities — from phenomenon observation, problem formulation, hypothesis construction, and variable identification, through to experimental planning, data collection, data analysis, and conclusion drawing — providing clear procedural guidance that minimized confusion during implementation. The integration of feedback components at strategic points within the LAM further enabled students to self-monitor their learning progress and make timely corrections, contributing to the smooth flow of learning activities as observed across all four meetings. These results are consistent with Hakim et al. (2024), who established that chemistry instructional materials grounded in real-world contexts substantially elevate student engagement and the perceived applicability of learning content.

Effectiveness of the LAM

Pretest and Posttest Results

Effectiveness of the developed LAM was determined based on the comparison of pretest and posttest scores of 34 students following the limited trial across four learning meetings. It is important to note that the results reported here reflect improvement observed after the limited trial within a one-group design, and should not be interpreted as definitive causal evidence of the LAM's superiority compared with other learning materials, given the absence of a control group. The descriptive statistics of pretest and posttest scores are presented in Table 17.

Table 167. Descriptive Statistics of Pretest and Posttest Scores

Statistics	Pretest	Posttest
Mean	40,20	85,66
Minimum Score	25,00	75,00
Maximum Score	75,00	95,83
Standard Deviation	13,37	5,45
N	34	34

The mean score increased substantially from 40.20 at pretest to 85.66 at posttest, representing a mean gain of 45.46 points. The Wilcoxon Signed-Rank Test revealed that all 34 students (100%) showed positive ranks, with no negative ranks or ties observed, yielding $Z = -5.093$ and $p < .001$. H_0 was therefore rejected, confirming a statistically significant difference between pretest and posttest scores. The effect size calculation further yielded $r = 0.87$, falling in the large category ($r \geq 0.50$), indicating that the improvement was not only statistically significant but also practically meaningful in magnitude.

To provide a more detailed profile of improvement across each science process skill indicator, pretest and posttest scores were analyzed per indicator alongside N-gain scores. The results are presented in Table 18.

Table 178. Pretest, Posttest, and N-Gain Score per Science Process Skill Indicator

No	Indicator	Pretest (%)	Posttest (%)	Gain (%)	N-Gain Score	Category
1	Formulating Problems	37,5	87,5	50	0,74	High
2	Constructing Hypotheses	45	77,5	32,5	0,61	Medium
3	Identifying Variables	42,5	87,5	45	0,81	High
4	Collecting Data	40	90	50	0,83	High
5	Analyzing Data	40	85	45	0,78	High
6	Drawing Conclusions	35	87,5	52,5	0,63	Medium
	Mean	40	85,83	45,83	0,73	High

As shown in Table 19, all six science process skill indicators demonstrated meaningful improvement following LAM implementation. Four indicators fell in the high N-gain category: collecting data (N-gain = 0.83), identifying variables (N-gain = 0.81), analyzing data (N-gain = 0.78), and formulating problems (N-gain = 0.74). Two indicators — constructing hypotheses (N-gain = 0.61) and drawing conclusions (N-gain = 0.63) — fell in the medium category, suggesting these indicators require more sustained scaffolding to achieve mastery. The relatively lower N-gain for constructing hypotheses is consistent with Azmi et al. (2021), who identified hypothesis formulation as one of the most cognitively demanding science process skills for secondary school students. Meanwhile, the medium N-gain for drawing conclusions despite its highest absolute gain (+52.50%) reflects its lower pretest baseline (35.00%), which provided greater room for raw score improvement but was offset by a lower starting point relative to the maximum score.

The overall mean N-gain of 0.73 (high category), combined with the large effect size ($r = 0.87$) and statistically significant Wilcoxon test result ($p < .001$), collectively indicate that LAM implementation was associated with substantial improvement in students' science process skills across all measured indicators. The integration of AfL strategic steps — particularly embedded feedback prompts, peer learning activities, and self-assessment components — within the guided inquiry LAM appears to have effectively supported students in developing science process skills across all four reaction rate sub-topics, consistent with Utami and Agustini (2024). Nevertheless, as this study employed a one-group design without a control group, the observed improvements cannot be exclusively attributed to the LAM; other contributing factors such as teacher facilitation quality, student motivation, and prior learning experiences should also be acknowledged.

CONCLUSION

This study successfully developed Student Activity Sheets (*Lembar Aktivitas Murid*/LAM) based on Assessment for Learning (AfL) principles for reaction rate material, specifically the sub-topic of factors affecting reaction rate, aimed at improving the science process skills of grade XI students at SMAN 1 Menganti. Based on the limited trial conducted with 34 students, the developed LAM demonstrated feasibility across three established criteria.

First, regarding validity, the LAM met the validity standard with a mode score of 4 (valid) across content, construct, and language validity dimensions, indicating that the material is scientifically accurate, structurally coherent, and linguistically appropriate for the target learners.

Second, regarding practicality, the LAM was rated very practical within the limited trial context, with a student response score of 98.53% and an activity observation score of 89.25%. The observation scores demonstrated a consistent upward trend across four meetings (84.00% to 95.00%), suggesting that students became increasingly familiar with the AfL-based learning procedures as the trial progressed.

Third, regarding effectiveness, the limited trial showed a statistically significant improvement in students' science process skill scores following LAM implementation, with mean scores increasing from 40.20 at pretest to 85.66 at posttest.

The Wilcoxon Signed-Rank Test confirmed this improvement ($Z = -5.093$, $p < .001$), supported by a large effect size ($r = 0.87$) and an overall mean N-gain of 0.73 (high category). These results indicate that the AfL-based LAM was associated with meaningful improvement in science process skills within the limited trial setting, though caution is warranted in generalizing these findings beyond the present research context given the one-group design and single-class sample.

Based on the limited trial, the AfL-based LAM was valid, very practical, and showed significant improvement in students' science process skill scores. The integration of AfL principles — particularly structured formative feedback, peer learning activities, and guided self-assessment — within a guided inquiry LAM provides a promising and empirically supported approach to developing science process skills in secondary chemistry education. Future studies are encouraged to replicate this work with larger and more diverse samples using a control group design to establish more robust evidence of the LAM's effectiveness.

RECOMMENDATION

Based on the findings and limitations of this study, the following recommendations are directed toward three audiences: teachers, future researchers, and product developers.

For Teachers

Teachers who wish to implement AfL-based LAM in their chemistry classrooms are advised to ensure sufficient time allocation for each learning session, particularly for activities requiring higher-order scientific reasoning such as hypothesis construction, experimental planning, and data analysis. Structured and individualized AfL feedback should be delivered at each stage of the guided inquiry process rather than only at the end of the session, as timely feedback is a critical mechanism through which AfL supports science process skill development. Teachers are also encouraged to conduct a brief orientation session prior to the first LAM meeting to familiarize students with the AfL procedural framework, thereby minimizing the adjustment period observed in the early meetings of this study.

For Future Researchers

Future researchers are advised to employ a quasi-experimental design with a control group to allow more robust causal inferences about the effectiveness of the AfL-based LAM compared with conventional learning materials. The sample size should be expanded beyond the 34 students used in this limited trial, and the study should be replicated across multiple classes and schools with diverse demographic and academic backgrounds to strengthen the generalizability of the findings. Additionally, future studies are encouraged to extend the trial period beyond four meetings to allow students sufficient time to fully develop indicators requiring sustained higher-order reasoning, particularly constructing hypotheses and drawing conclusions, which demonstrated medium-category N-gain scores in the present study. Researchers may also explore the application of AfL-based LAM across different chemistry topics beyond reaction rate — such as chemical equilibrium, electrochemistry, or thermochemistry — to examine the broader applicability of this instructional approach.

For Product Developers

Product developers and instructional designers are encouraged to build upon the LAM framework developed in this study by incorporating dynamic and adaptive feedback mechanisms that can respond to individual student responses in real time, addressing the limitation of fixed pre-written feedback prompts identified in the present study. The LAM design could also be extended to a digital or interactive format to facilitate more flexible and personalized AfL feedback delivery. Furthermore, the number of LAM units may be expanded to cover additional reaction rate sub-topics or integrated into a broader chemistry learning module to support more comprehensive science process skill development across an extended learning sequence.

BIBLIOGRAPHY

Aldiyah, E. (2021). Lembar Kerja Peserta Didik (LKPD) pengembangan sebagai sarana peningkatan keterampilan proses pembelajaran IPA di SMP. *TEACHING: Jurnal Inovasi Keguruan Dan Ilmu Pendidikan*, *1*(1), 67–76.

Arviani, V. (2011). *Identifikasi pemahaman konsep laju reaksi siswa kelas XI SMA Brawijaya smart school Malang*.

Assessment Reform Group. (2002). Assessment for learning: 10 principles: Research-based principles to guide classroom practice. *University of Cambridge, School of Education*. https://www.assessment-reform-group.org

Azmi, N. L., Nurhayati, S., Wardani, S., & Priatmoko, S. (2021). Pengembangan instrumen tes untuk mengukur HOTS peserta didik pada materi laju reaksi. *Chemistry in Education*, *10*(1), 45–52.

Basit, D. A., Muslim, B., & Saridewi, N. (2023). Pengaruh model problem based learning berbasis etnosains terhadap hasil belajar siswa pada materi laju reaksi. *Spin Jurnal Kimia & Pendidikan Kimia*, *5*(1), 75–90.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Budiyono. (2011). *Penilaian Hasil Belajar*. Universitas Sebelas Maret Press.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. routledge.

Fitrianna, A. Y., & Anita, I. W. (2017). Pengembangan LKS Berbasis Assessment for Learning (AfL) Teman Sejawat untuk Kemampuan Matematis Siswa SMP. *JURNAL E-DuMath*, *3*(2). https://doi.org/10.52657/je.v3i2.454

Gürses, A., Çetinkaya, S., Doğar, Ç., & Şahin, E. (2015). Determination of Levels of Use of Basic Process Skills of High School Students. *Procedia - Social and Behavioral Sciences*, *191*, 644–650. https://doi.org/10.1016/j.sbspro.2015.04.243

Hake, R. R. (1999). *Analyzing Change/Gain Scores*. Indiana University. https://www.physics.indiana.edu/~sdi/AnalyzingChange-Gain.pdf

Hakim, E., Astafani, A., & Resmawati, R. F. (2024). Systematic Review Faktor-Faktor Kesulitan Belajar Materi Kimia. *Jurnal Inovasi Pendidikan Kimia*, *18*(2), 81–88.

Lutfi, A. (2021). Research and Development (R&D): Implikasi dalam pendidikan kimia. *Surabaya: Jurusan Kimia FMIPA Universitas Negeri Surabaya*.

Maskur, M. (2023). Dampak Pergantian Kurikulum Pendidikan terhadap Peserta Didik Sekolah Dasar. *Jurnal Keguruan Dan Ilmu Pendidikan (JKIP)*, *1*(3), 190–203. https://doi.org/10.61116/jkip.v1i3.172

Maulani, G., Septiani, S., Susilowaty, N., Rusmayani, N. G. A. L., Evenddy, S. S., Nababan, H. S., Setiadi, K., Rahayu, I., Simanungkalit, L. N., & Edi, S. (2024). *Evaluasi Pembelajaran*. Sada Kurnia Pustaka.

Nurfidayanti, H. N., & Yonata, B. (2022). Development of student worksheets based on guided inquiries to train students science process skills on reaction rate materials. *Jurnal Pijar Mipa*, *17*(1), 8–15.

Perubahan atas Keputusan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 56/M/2022 tentang Pedoman Penerapan Kurikulum dalam Rangka Pemulihan Pembelajaran, 262/M/2022 Keputusan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi (2022). https://jdih.kemdikbud.go.id

Plomp, T., & Nieveen, N. (2013). *Educational design research – Part A: An introduction*. SLO. http://international.slo.nl/edr

Prastowo, A. (2015). *Panduan kreatif membuat bahan ajar inovatif*. Diva Press.

Pratomo, H., Fitriyana, N., Wiyarsi, A., & Marfuatun, M. (2025). Mapping chemistry learning difficulties of secondary school students: A cross-grade study. *Journal of Education and Learning (EduLearn)*, *19*(2), 909–920.

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 21–33.

Riduwan. (2015). *Skala Pengukuran Variabel-Variabel Penelitian*. Bandung: Alfabeta.

Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D* (2nd ed.). Bandung: Alfabeta.

Tawil, Muh., & Liliasari. (2014). *Keterampilan-keterampilan sains dan implementasinya dalam pembelajaran IPA* (Cetakan 1). Universitas Negeri Makassar.

Thiagarajan, S., Semmel, S. S., & Semmel, M. I. (1974). *Instructional Development for Training Teachers of Exceptional Children: A Sourcebook*. Council for Exceptional Children. https://eric.ed.gov/?id=ED090725

Utami, D. F., & Agustini, R. (2024). Development of Student Worksheets based Assessment for Learning on Chemical Equilibrium to Improve Student Science Process Skills. *Hydrogen: Jurnal Kependidikan Kimia*, *12*(4), 834. https://doi.org/10.33394/hjkk.v12i4.12617