



Implementing Peer Assessment in Introductory Chemistry Labs: Educational Impact and Student Perceptions

Kenichi Goto

Tokyo University, Japan

Abstract

This study investigates the educational impact of peer assessment in laboratory report writing instruction for first-year university students in Japan. Given the national concern over insufficient training in evidence-based thinking and academic writing [1], the study targeted a foundational chemistry lab course involving 44 students with varying prior experience. Over 15 weeks, three peer assessment sessions were implemented, and pre- and post-surveys were conducted to evaluate changes in student attitudes [2].

Findings revealed that students without prior experience gained confidence and motivation, citing peer feedback as instrumental in understanding how to improve their writing. They also reported growth in scientific expression. Conversely, experienced students became more aware of content quality and used peer comparisons to refine their work [3]. Across both groups, peer assessment fostered self-regulated learning and deeper engagement with the writing process [4].

The results suggest that peer assessment is an effective pedagogical strategy for developing academic writing and critical thinking skills in STEM education. By positioning students as active evaluators, it promotes reflective learning and supports essential competencies for academic success [5]. The study emphasizes the value of structured feedback and collaborative learning in bridging the educational gap between secondary and higher education.

Keywords: Peer Assessment, Academic Writing, Introductory Chemistry Laboratory

1. Purpose of the study

Many scholars have pointed out the need for instruction in laboratory reports, which are indispensable for university students' learning (e.g., Suzuki, 2009) [6]. Therefore, this study surveyed first-year students enrolled in a science-oriented department (Department of Food and Environmental Science) at University A in Japan regarding chemistry experiments and report writing (2025; $n=132$). The results showed that the subjects in which students had conducted the most experiments in high school were chemistry (23.8%), basic chemistry (17.2%), biology (17.2%), basic biology (4.9%), and basic physics (4.1%), while other subjects were below 4%.

Regarding the frequency of chemistry experiments, the largest proportion (25.4%) reported "never" having done them, followed by "once per semester" (23.0%), "once per month" (20.5%), and "rarely" (13.9%). Concerning chemistry experiment reports, only 11.5% of students had experience preparing reports separate from the laboratory handouts used in class; 46.7% had submitted the class laboratory handouts, and 41.8% had no experience of submission. Furthermore, when asked whether they felt anxious about writing chemistry experiment reports at university, 77% of students reported feeling anxious. The main sources of anxiety were: writing and structure (62%), discussion/interpretation (49%), lack of experience (38%), creating figures and graphs (30%), length and grading (19%), and Japanese writing ability (11%).

2. Definition and Theoretical Background of Peer Assessment

In this study, peer assessment refers to an educational practice in which learners evaluate one another's work, thereby embedding assessment activities within the learning process to promote learners' self-regulation and autonomous learning. This approach aligns with Earl's concept of Assessment as Learning, which reconceptualizes assessment from a teacher-driven means of assigning grades to an activity through which learners themselves regulate and advance their learning [7]. Concretely, the method emphasizes sharing clear assessment criteria (rubrics) based on learning objectives, having learners first conduct self-assessment, then receive anonymized peer assessments



from multiple evaluators, and finally use that feedback to revise and re-evaluate their work—forming a cyclical process of self-assessment → peer assessment → re-self-assessment [8].

A series of studies by Gotō (Gotō et al.) demonstrates that practices using peer-assessment forms contribute to improvements in students' ability to construct discussions, their scientific expression skills, and their sense of self-efficacy in the context of chemistry laboratory work; these studies empirically support the synergistic effects of combining self- and peer-assessment[[9]. Moreover, beyond traditional face-to-face, paper-based implementations, the use of online platforms has been reported to enable peer assessment that overcomes temporal and spatial constraints; features such as commenting and rating functions facilitate real-time interaction and deepen learning[10]. Through these processes, learners gain opportunities to incorporate others' perspectives to objectively review their own descriptions and interpretations, and to experience the self-regulatory processes that assessment can stimulate, thereby promoting consolidation of learning.

At the same time, variability among raters and differences in the quality of feedback remain persistent challenges; ensuring reliability therefore requires rater training, instructor moderation, and clear specification of assessment criteria[11]. Accordingly, this study incorporates explicit rubrics, rater training, anonymization and multiple raters, and random instructor re-evaluation (moderation) into the design, and examines changes in “understanding of writing and structure,” “level of anxiety,” and “confidence” through pre- and post-questionnaires and qualitative analysis of reflections. Overall, peer assessment is an effective means of fostering critical thinking, collaborative skills, and self-regulation; by drawing on these theoretical and practical insights, the present study introduces peer assessment to improve learning outcomes and reduce anxiety in the preparation of chemistry laboratory reports.

3. Research Methods

Participants and Sampling

Participants: The preliminary survey targeted 132 first-year students in the Department of Food and Environmental Sciences at University A.

The subsequent survey targeted 43 students enrolled in the author's course.

Sampling Method:

A census approach was adopted, targeting all students enrolled in the author's course.

Instructional Materials and Assessment Tools

Experimental Materials:

Six chemistry experiments were conducted, selected to allow comparison based on the objectives of each session. The experiments included the following basic laboratory activities:

Basic Analytical Chemistry: 1 Acid–base titration (quantification of acetic acid in vinegar), 2 Buffer solutions and pH, 3 Determination of copper by spectrophotometry, 4 Determination of protein by the Biuret method

Applied and Everyday Chemistry Experiments: 5 Ethanol production using immobilized yeast (two consecutive weeks) 6 Synthesis and purity testing of aspirin (two consecutive weeks)

Peer assessment was conducted after each experiment.

Peer Assessment Activities: A peer-assessment rubric was used for each experiment, consisting of five performance levels with specific descriptors aligned with the experiment's objectives, methods, results, discussion, and challenges.

The evaluation sheet used the same rubric and included sections for self-assessment scores and written comments.

Assessment Criteria:

The following elements were evaluated: Writing quality and structure (organization of purpose, methods, results, and discussion)

Logical quality of the discussion (hypothesis formulation, data interpretation, validity of conclusions)

Appropriateness of figures and tables (choice of graphs/tables, annotations, readability)

Expression (clarity of Japanese writing, use of technical terms, conciseness)

Significant figures, calculations, and references

Questionnaire Surveys

Questionnaires regarding the chemistry experiments and peer-assessment activities were administered after the 1st and 6th sessions (using a four-point scale and open comments).

Additional items were included only in the final (6th) survey.



Procedures

1. Preparation: The research purpose, procedures, and data usage were explained, and informed consent was obtained.
Rater training was conducted, including rubric explanation and mock evaluation exercises (instructor model evaluation → student evaluation → instructor feedback).
2. Laboratory Sessions and Report Writing:
Regular laboratory classes were conducted, and students prepared laboratory reports.
3. Self-Assessment:
Before submission, students completed a self-assessment sheet and wrote self-reflection comments.
4. Peer Assessment:
Reports were randomly assigned and evaluated using paper-based evaluation sheets.
Raters assigned scores and improvement comments based on the rubric.
5. Feedback Distribution:
Each student received aggregated peer-assessment results and comments.
6. Reflection and Revision:
Students prepared improvement plans based on the feedback and submitted a short revision statement.
7. Re-evaluation:
The instructor evaluated the revised statements to confirm learning changes.
8. Questionnaire Surveys:
Surveys on the experiments and peer assessment were administered in the 1st and 6th sessions.
9. Data Collection:
Quantitative data: rubric scores, pre/post questionnaire scores, submission rates
Qualitative data: reflection statements, peer-assessment comments
10. Data Analysis:
Normality was tested, and pre–post comparisons were conducted using paired t-tests or Wilcoxon signed-rank tests.
Effect sizes (Cohen's d or r) were calculated ($r = Z/N$) and interpreted based on Cohen's criteria.
Inter-rater reliability was calculated using Cohen's κ or ICC, and the rubric was revised as needed.
Quantitative and qualitative data were integrated to examine changes in the learning process and mechanisms of anxiety reduction.
11. Post-analysis Tools:
Rubrics, evaluation sheets, and analysis software (SPSS, KH Coder, etc.) were used.

The open-ended responses and reflection entries were analyzed qualitatively using the following procedure. First, the author closely read all data and extracted initial codes through open coding. The extracted codes were then organized into axial codes to construct major themes, and a growth model was defined through selective coding. Coding was performed independently by two coders; interrater agreement was calculated for the initial 50 cases (Cohen's $\kappa = 0.78$). Discrepancies were resolved through discussion, and a final codebook was produced. The analysis combined automated text-mining with KH Coder (version 3.BETA0.5) and manual thematic review.

4. Results and discussion

4.1 Summary of Questionnaire Responses (point ratings)

A paired Wilcoxon signed-rank test was conducted on four-point scale items before and after the introduction of peer assessment ($N = 44$). **Report comprehension.** Significant improvements were observed for “**What should be written in the lab report**” ($Z = 2.52$, $p = .012$, $r = .38$) and “**How to write it**” ($Z = 2.29$, $p = .022$, $r = .35$). These results indicate a moderate effect of peer assessment on students' understanding of report structure and writing style.

Assessment behavior. Significant increases were found for “**Was self-assessment easy?**” ($Z = 2.49$, $p = .013$, $r = .38$) and “**Was evaluating others' reports easy?**” ($Z = 2.12$, $p = .034$, $r = .32$). Students initially reported limited assessment experience and difficulty, but repeated practice appears to have improved perceived ease of assessment.

Affective aspects. The item “**I enjoyed doing self-assessment**” increased significantly ($Z = 2.00$, $p = .045$, $r = .30$), suggesting that peer assessment may have enhanced positive feelings toward self-evaluation.



Non-significant items. “Peer assessment proceeded smoothly” ($p = .086$) and “I enjoyed assessing others” ($p = .089$) did not reach significance. Although trends were present, large individual differences in enjoyment and smoothness suggest a need for additional support and time allocation. “Will my evaluation help improve my future writing?” showed no difference ($p = .972$), indicating that self-assessment does not necessarily translate directly into revision behavior.

Understanding, confidence, and grades. Large effect-size improvements were found for “Understanding what to write in a chemistry lab report” ($Z = -3.64$, $p < .001$, $r = .55$) and “Confidence in writing reports” ($Z = -4.78$, $p < .001$, $r = .72$). Report grades were significantly correlated with both understanding and confidence ($r = .56-.72$). These findings suggest that the educational interventions, including peer assessment, strongly contributed to deeper understanding and confidence building.

Perceived value of assessment. Items addressing the value of point-based and comment-based evaluation showed no significant changes ($p = .537-.827$), likely because many students already recognized the value of assessment activities at the first measurement and maintained that positive attitude. Below Table 1 is a summary table of the point-rating results.

Table 1

Item	N	Z	p (two-tailed)	$r = Z/\sqrt{N}$	Interpretation
What to write in a lab report (1 vs 6)	44	2.524	.012*	.381	$p < .05$
How to write a lab report (1 vs 6)	44	2.294	.022*	.346	$p < .05$
Peer assessment proceeded smoothly (1 vs 6)	44	1.719	.086	.259	n.s.; $p = .086$
I enjoyed doing self-assessment (1 vs 6)	44	2.001	.045*	.302	$p < .05$
I enjoyed assessing others (1 vs 6)	44	1.699	.089	.256	n.s.; $p = .089$
Was evaluating my own report easy? (1 vs 6)	44	2.490	.013*	.375	$p < .05$
Was evaluating others' reports easy? (1 vs 6)	44	2.118	.034*	.319	$p < .05$
Will my evaluation help improve my future writing? (1 vs 6)	44	-0.035	.972	-.005	n.s.; $p = .097$
Felt point evaluation of my report was meaningful	44	-0.617	.537	.09	n.s.; $p = .054$
Felt comment evaluation of my report was meaningful	44	-0.428	.669	.06	n.s.; $p = .064$
Felt point evaluation of others' reports was meaningful	44	0.218	.827	.03	n.s.; $p = .083$
Felt comment evaluation of others' reports was meaningful	44	-0.409	.682	.06	n.s.; $p = .068$
Understanding vs Confidence	44	-3.640	$p < .001$.55	Statistically significant ($p < .001$); large effect ($r = .55$)
Report grade vs understanding what to write	44	-3.712	$p < .001$.56	Statistically significant ($p < .001$); large effect ($r = .56$)
Report grade vs confidence in writing reports	44	-4.780	$p < .001$.72	Statistically significant ($p < .001$); large effect ($r = .72$)

Note: Paired two-group (pre/post) Wilcoxon signed-rank tests (two-tailed) were used as described in Methods. Significance level $\alpha = .05$.

$r = Z/\sqrt{N}$, * $p < .05$; ** $p < .01$



4.2 Summary of Open-Ended Responses

Open-ended questionnaires about attitudes toward lab reports were collected at the 1st and 6th sessions. Comparison between the first and sixth responses revealed the following.

Table 2: Overall pattern of growth

Aspect	Characteristics at 1st session	Characteristics at 6th session	Key change
Clarity of goals	Abstract goals (e.g., “want to write smoothly,” “write carefully”)	Concrete goals (e.g., “attend to significant figures,” “draw structural formulas,” “title placement for figures”)	Shift from abstract to concrete; clearer action-level intentions
Understanding of report structure	Beginning to notice format (purpose, results, etc.)	Understood flow: purpose → principle → method → results → discussion → conclusion	Grasp of report “template” and logical structure
Scientific thinking	“Want to understand and write”	Consideration of “why these results,” literature comparison, causal relations	Move from descriptive writing to analytical thinking
Learning attitude	“Want to prepare and review; avoid last-minute work”	“Reflect immediately after experiment,” “allow time,” “acquire specialized knowledge”	Concrete behavioral improvements and increased autonomy
Expression and presentation	“Want to use tables and graphs”	Attention to spacing for units, figure titles, use of white space	Improved presentation skills and reader awareness
Originality	Referencing senior reports	“Use my own words,” “originality,” “write independently”	Transition from imitation to independence

Table 3: Changes in keywords (vocabulary development)

Domain	Frequent words at 1st	Words increased by 6th	Meaning of change
Effort and attitude	smooth, careful, prep/review, nitpicky, last-minute	margin, review, record, specialized, deeper understanding	Improved quality of actions
Report technique	table, graph, readability	significant figures, structural formula, reaction equation, title placement, units	Increase in technical and domain-specific vocabulary
Depth of thought	easy to understand, own words	reason, literature comparison, significance, cause, insight	Acquisition of scientific thinking
Autonomy	refer to seniors’ reports	write independently, originality, ingenuity	Growth toward self-directed learner

Table 4: Growth stage model

Level	State at 1st	State at 6th	Interpretation
Level 1 Formal understanding	Beginning to know what to write	Understands basic report structure	Foundation established
Level 2 Technical skill acquisition	Wants to use tables/graphs	Attends to formatting, units, titles	Refinement of technique
Level 3 Scientific thinking	Wants clarity	Considers reasons, literature, significance	Deepening of thought
Level 4 Autonomous learning	Uses seniors’ reports as reference	Writes independently with originality	Emergence of independent learner
Level 5	Wants to write	Records observations, notes	Seed of research



Level	State at 1st	State at 6th	Interpretation
Researcher-like attitude	carefully	insights, analyzes causes	mindset

4.3 Outcomes and Challenges of Introducing Peer Assessment

Based on the study objectives, the outcomes and challenges of implementing peer assessment are summarized below.

Deeper understanding of report structure

1. **Outcome.** Through peer assessment, students referenced others' report structures, explanatory approaches, and figure usage to identify specific deficiencies in their own reports. Common comments included: *"I noticed omissions I wouldn't have seen," "I learned structural and expressive techniques," "I understood what kind of report to write,"* and *"I realized intermediate steps needed explanation."* This indicates that peer assessment functioned as a learning resource for understanding the basic report structure (purpose, method, results, discussion).

Challenge. Some students reported that assessment criteria were ambiguous, making evaluation difficult and inconsistent. Improved rubric development is needed to support structural understanding.

2. **Enhanced self-understanding and self-assessment**

Outcome. Comparing their work with peers helped students clarify strengths and weaknesses: *"Good and bad points became clear," "I could analyze my weaknesses," "I confirmed that my writing approach was appropriate."* Peer assessment thus acted as a metacognitive activity that improved self-assessment accuracy.

Challenge. Issues included leniency in peer ratings, biased evaluations among acquaintances, and difficulty maintaining objectivity. Mechanisms to ensure objectivity (anonymity, random assignment) are necessary.

3. **Promotion of peer learning**

Outcome. Peer assessment was valued as an opportunity to learn others' techniques, explanations, and presentation skills: *"I learned ideas I didn't have," "I found points I want to emulate," "I can apply others' strengths next time."* This suggests peer assessment supported the class as a learning community and broadened learning opportunities.

Challenge. Students reported difficulty transferring others' strengths into their own writing and limited time to review peers' reports thoroughly. Instructional design to support transfer is required.

4. **Anxiety reduction and increased motivation**

Outcome. Peer assessment reduced anxiety about report writing and increased motivation to improve: *"I wanted to write a report I'm not embarrassed by," "I felt motivated to try harder," "I want to apply this next time."* Peer assessment positively affected learners' psychological states.

Critical challenge. Even students who valued peer assessment did not always gain confidence from it: *"I can't say it led to confidence," "I worry about rating differences."* Peer assessment alone may be insufficient to raise self-efficacy; additional supports are needed.

5. Conclusions and Future Issues

This study examined the effects of peer assessment on students' laboratory report writing. Reading peers' reports helped students detect structural omissions and insufficient explanations that they might have missed, and tended to clarify the basic report structure (purpose, method, results, discussion). Referring to others' expression styles and figure usage provided concrete cues for improving one's own writing, suggesting that peer assessment contributes to deeper understanding of report composition. The experience of being evaluated by others also prompted students to view their writing more objectively and to adopt a reader-oriented perspective.

Peer assessment also had a measurable effect on reducing anxiety about report writing. By comparing their work with others, students identified strengths and areas for improvement and developed concrete plans for future revisions. However, the study revealed that even high-achieving students or those who recognized the value of peer assessment did not necessarily gain confidence from the process. Ambiguous evaluation criteria, leniency among acquaintances, and difficulty internalizing others' strengths were identified as factors that can hinder confidence formation. These issues must be addressed to maximize the educational benefits of peer assessment.



In light of these findings, peer assessment is effective for promoting understanding of report structure and reducing anxiety, but additional support is required to foster students' confidence. Practical improvements include clarifying evaluation criteria, ensuring anonymity, and providing instruction to help students transfer peers' strengths into their own writing. Future work should systematize the peer-assessment process, create a safe environment for reciprocal evaluation, and develop interventions to enhance self-efficacy in report writing. Such measures would help peer assessment function not merely as an evaluative activity but as a sustainable mechanism that supports ongoing student learning.

REFERENCES

- [1] Tanaka, K. (2021). Fostering Evidence-Based Thinking and the Transition from Secondary to Higher Education. *Journal of Educational Evaluation Research*, 20, 45–58.
- [2] Suzuki, T. (2009). *Introduction to Academic Writing for University Students*. Tokyo: Gakujutsu Shuppankai.
- [3] Goto, K. (2013). Fostering Scientific Expression through Peer Assessment Activities. *Journal of Science Education Research*, 54, 32–41.
- [4] Iida, S., & Goto, K. (2019). Practice and Challenges of Peer Assessment Activities in Junior High School Science Classes. *Journal of Educational Practice Research*, 8, 15–26.
- [5] Earl, L.M. (2003). *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning*. Thousand Oaks, CA: Corwin Press.
- [6] Hiroaki Suzuki (2009). *Writing Skills Generated by Collaborative Learning: An Attempt at Report-Writing Education at University*. Maruzen Planet.
- [7] L.M. Earl, *Assessment as Learning*, C.A. Corwin Press, 2003.
- [8] Norihiko Kitao (Kitō), *Zuwode Wakaru Kyōshoku Skill Up Series 3: Manabi o Hikidasu Gakushū Hyōka*, Zoshobunka, 2006.
- [9] Kenichi Goto, (2013) "A Study on the Effects of Self-Assessment in High School Chemistry Experiments: Utilizing Peer-Assessment Forms," *Research in Science Education*, Vol.54, No.1, pp.13–26,
- [10] Nasu (2014), related report on learner motivation and online assessment.
- [11] Azuma (2001), Tanaka (2013), Abiko (1987): considerations on assessment theory and practice.