Challenges in assessing chemistry lab reports among pre-service teachers

Subject and problem statement

Chemistry labs provide students with a distinctive learning experience. The beginning of the 21st century marked a change in chemistry education via the introduction of the inquiry-based chemistry labs (Cacciatore & Sevian, 2009; Wheeler et al., 2017). Inquiry enhances content understanding via the connection between theory and observed data (Chiu & Linn, 2014) and promotes skills like asking research questions, making hypothesis, providing scientific explanations, and communicating arguments (van Brederode et al., 2020; Walker & Sampson, 2013). The importance of learning chemistry by performing experiments is clear in supporting students' understanding of submicroscopic thinking and its connection to the macroscopic level (Authors, 2015). The process of evaluating an inquiry activity is essential for both students and teachers as it influences the teaching and learning process (Huang & He, 2016). Using rubrics not only make teacher's requirements explicit, but also, they provide students with clear expectations at a high-performance level (Authors & colleagues, 2019). However, using a rubric can be challenging; designing a rubric in a coherent manner is a complex task. Moreover, consistency among graders is essential to maintain accuracy in providing students with improvement direction (Allen, & Tanner, 2006; Author & colleagues, 2019). Studies on professional development (PD) of high-school chemistry teachers, as well as pre-service chemistry teachers on assessing chemistry lab reports using rubrics, are limited. Some studies that have been conducted at the undergraduate level, documenting Teaching Assistants (TAs) while assessing students in the lab, concluded that most TAs had difficulty in providing concise and accurate written feedback (Authors & colleagues, 2019; Kurdziel et al., 2003). For high school chemistry students to develop understanding as well as scientific skills through inquiry-based chemistry labs, their teacher's lab-report assessment and feedback is crucial. Thus, teachers' PD and teachers' reflections on assessing inquiry-based chemistry lab reports is important. It can provide them the opportunity to: (a) better understand how and why rubrics are constructed, and (b) encourage teachers' reflective practice of the inquiry process while providing meaningful, valuable, and consistent feedback that enhances students' learning. The goal of the research was to understand how explicit guidance of assessing inquiry-based lab reports affected pre-service chemistry teachers' lab score variation and written feedback to students. Towards that end three research questions were posed: **A**) What was the effect of the explicit guidance on lab score variation on 1st assessment vs. the 2nd assessment? **B**) What was the effect of the explicit guidance on pre-service teachers' written feedback given to students' lab reports on 1st assessment vs. the 2nd assessment? **C**) What was the perception of the pre-service teachers on the process of lab reports' assessment?

Design and procedure

Chemistry inquiry labs had been an important component of the high-school curriculum and the matriculation exam in our country. To implement inquiry-based chemistry labs, teachers were guided through professional development programs, and an assessment tool was developed. The assessment tool was a rubric which contained a section for assessing the different phases of an inquiry-type experiment as well as a section for teacher's observations of the student's group work in the lab (Hofstein et al., 2004). During 11th and 12th grades, every chemistry-major student must conduct 8 different experiments and generate a report for each one based on a rubric provided by the Chemical division of the ministry of Education. The report is assessed by the teacher in a formative style, then corrected by the student and resubmitted for an additional assessment. Finally, all reports are organized in a portfolio, and an oral exam is administered.

Research participants. Thirteen pre-service chemistry teachers, who enrolled to teachers' preparation program, participated in the research. This program spans over two years and covers topics in general education and chemistry education. All the participants volunteered to participate in the research and signed a consent form.

Research methodology. A mixed methods approach was used, combining both qualitative and quantitative tools as method of analysis, to provide better inferences and minimizing unimethod bias (Tashakkori & Teddlie, 2015). This specific research was conducted during a course taught on the topic of teaching high-school inquiry-based chemistry labs. The course included theoretical aspects of what inquiry-based chemistry labs are, the different levels of inquiry-based experiments, and the assessment process and methods of lab reports. The first stage of the research was dedicated to explanation on using the rubric properly – what each criterion means and what a proper response to a criterion is – with only relatively few instructions and comments on how to provide a written feedback. In the 2nd step, the pre-service teachers assessed the first anonymous lab report – scored it based on the rubric and provided a written feedback in the body of the report – and submitted their assignment for the professor's evaluation. In the 3rd stage a reflection was conducted on the 1st lab report assessment. The professor delivered an analysis of the 1st report and a discussion was held regarding their rubric grading, correctly using the rubric, the understanding of the meaning of each criterion, as well as their written feedback to students. The aim was to assist the pre-service teachers in improving score consistency of their 2nd lab report assessment which they had to submit in the final step, along with a response to a reflective questionnaire about the entire lab report assessment process.

Research tools. The three research tools used were aligned with the three research questions: **A**) Quantitative analysis of Rubric score provided by pre-service teachers on two lab reports. A data-table was created and then analyzed. **B**) Qualitative content analysis of the written feedback given to students by pre-service teachers on two lab report, based on four categories that emerged from the data (Hsieh & Shannon, 2005) . **C**) Qualitative content analysis of the reflection written by pre-service teachers, based on six categories that emerged from the data. Two researchers discussed the analysis and cross-checked the data analysis to ensure trustworthiness (Guba, 1981).

Findings and Analysis

The inquiry-based lab report assessment was comprised of two components – scoring based on the rubric and providing written feedback to students in the body of the report. The rubric is highly comprehensive and is broken up into 3 sections. There are total of 10 dimensions listed under those sections, and anywhere between 1 to 6 different criteria under each dimension which make a total of 28 different criteria. A lab report must be assessed based on the rubric and each criterion must be scored by assigning it a whole number between 0 to 5. High-school students who write a lab report are given the rubric and they must structure the report based on it and follow its guidelines. The teacher should provide a written feedback by inserting comments in the body of the report and articulate what was missing or inaccurate, as well as, of course, complimenting for a job well done. In most cases, those comments should not simply state the correct answer but rather should be in a form of leading questions or provide direction to improve and promote further learning.

Overall rubric score. Overall score of the 1st assessment, provided by the 13 pre-service teachers, ranged between 40 & 90 and between 51 & 88 for the 2nd one. This large spread of the overall score for the two assessed reports was also expressed by the magnitude of the standard deviation values (13 for the 1st and 11 for the 2nd). The score spread and standard deviation were slightly better for the 2nd assessment but remained very high. In both assessments, the score of 10 out of 13 pre-service teachers ranged between 70-90, while the rest scored below 70 (potential outliers).

Table 1. Rubric score by section

Section	Available points	Measure	1st Assessment	2nd Assessment
1 - Getting familiar with a phenomenon	10	Average	7.1	8.1
		Std Dev	1.0	1.1
		Highest	8.7	10.0
		Lowest	5.3	6.7
		Range	3.3	3.3
2 - Experiment Planning	40	Average	32.8	31.7
		Std Dev	**5.9**	**3.2**
		Highest	40.0	37.0
		Lowest	15.0	24.5
		Range	**25.0**	**12.5**
3 - Experiment Execution	50	Average	36.1	34.3
		Std Dev	**6.9**	**8.4**
		Highest	43.6	45.0
		Lowest	20.0	20.0
		Range	**23.6**	**25.0**

Rubric score by section.
The rubric consisted of 3 major sections. The 1st one was 'getting familiar with a phenomenon' and its potential highest score was 10. The 2nd section was 'experiment planning' with a maximum score of 40, and the 3rd section was 'experiment execution' and its maximum score was 50. For each of the sections we calculated average, standard deviation, highest & lowest score, as well as range which are all shown in Table 1. In section 1 score standard deviation and range were consistent across the two assessments. In section 2 we see improvement from the 1st assessment to the 2nd assessment. The range was cut in half from 25 to 12.5 points, and standard deviation decreased by 47% from 5.9 to 3.2. In section 3, however, the 2nd assessment range and standard deviation had increased compared to the 1st one. Range slightly increased from 23.6 to 25, and standard deviation increased from 6.9 to 8.4. Next, dimensions score under sections 2 & 3 would be explored since they are the essence of the report.

*Table 2. **Section 2** score by dimension*

Dimension	Avail. points	Measure	1st Assessment	2nd Assessment
2.1 - Asking research questions	5	Std Dev	1.4	0.9
		Highest	5.0	5.0
		Lowest	0.0	2.0
		Range	5.0	3.0
2.2 - Formulating research question	10	Std Dev	**1.4**	**1.0**
		Highest	10.0	10.0
		Lowest	5.0	7.0
		Range	**5.0**	**3.0**
2.3 - Formulate a Hypothesis	10	Std Dev	**1.9**	**2.2**
		Highest	10.0	10.0
		Lowest	3.0	2.0
		Range	**7.0**	**8.0**
2.4 - Experiment plan	15	Std Dev	**2.7**	1.4
		Highest	15.0	14.0
		Lowest	5.0	9.5
		Range	**10.0**	**4.5**

Rubric score by dimension.
Section 2 'Experiment Planning' consisted of 4 dimensions. Dimensions data under section 2 is shown in Table 2. Across all dimensions, and for both assessments, we see relatively high range and standard deviations values. Range and standard deviation values were decreased from the 1st assessment to the 2nd assessment in dimensions 2.1, 2.2, & 2.4, especially for the dimension 'Experiment plan' where its standard deviation was cut by almost 50% from 2.7 to 1.4. However, range and standard deviation were increased in dimension 2.3 – 'Formulate a Hypothesis'. This dimension also exhibited large standard deviation values in both assessments.
Section 3 'Experiment Execution' consisted of 5 dimensions and the data for the dimensions under section 3 is shown in Table 3. Dimension 3.1 'Experiment Handling' was omitted since it could only be scored by observing students performing an experiment and therefore was artificially scored. Large standard deviation and range values were observed for dimensions 3.2, 3.3, & 3.4 in which students had to display & analyze the data, draw conclusions, as well as critically discuss

*Table 3. **Section 3** score by dimension*

Dimension	Avail. points	Measure	1st Assessment	2nd Assessment
3.1 - Experiment Handling	5	Std Dev	0.3	0.1
		Highest	5.0	5.0
		Lowest	4.0	4.5
		Range	1.0	0.5
3.2 - Results display and analysis	15	Std Dev	**2.2**	**2.6**
		Highest	14.3	12.8
		Lowest	6.0	3.0
		Range	**8.3**	**9.8**
3.3 - Draw conclusions	10	Std Dev	**2.4**	**3.0**
		Highest	10.0	10.0
		Lowest	3.0	0.0
		Range	**7.0**	**10.0**
3.4 - Summarized discussion	10	Std Dev	**1.8**	**2.5**
		Highest	9.3	8.0
		Lowest	2.0	0.7
		Range	**7.3**	**7.3**
3.5 - Overall lab report	10	Std Dev	1.8	1.3
		Highest	10.0	10.0
		Lowest	4.0	5.3
		Range	6.0	4.7

the results and the validity of the conclusions. In all these dimensions the standard deviation value had increased from the 1st to the 2nd assessment.

Lab report written feedback. In addition to scoring the two assessed lab reports based on the rubric, the 13 pre-service teachers were required to provide feedback in the body of the lab report in order to help and direct the students on how to improve the content of the report. The lab report was provided in a WORD document and feedback was inserted in the document itself in a form of comments and/or corrections of the actual text. Four different categories were developed to assess the quality of the feedback: Each assessed lab report was evaluated based on the categories shown in Table 4 and scored on a scale of Poor, Average, and Best. For each assessment, under each category, number of responses were summed by Poor, Average, and Best. Under all categories, improvement was observed on the 2nd assessment compared to the 1st one – more pre-service teachers obtained the scores 'Best' or 'Average' on their written feedback.

*Table 4. Pre-service teachers' - **written feedback** analysis*

Assessment Categories of Pre-service Teachers' Written feedback	Measure	1st Assessment	2nd Assessment
Category 1: Generally, provides constructive feedback (vs. correcting the text)	Poor	1	0
	Average	6	5
	Best	6	8
Category 2: Provides specific chemistry related feedback (including examples) as well as non-chemistry related feedback	Poor	1	0
	Average	6	8
	Best	6	5
Category 3: Provides feedback where student is required to consult or search for answers (vs. giving the answers). Especially under analysis, conclusions, and discussion dimensions under section 3 (Dimensions 3.2, 3.3, & 3.4).	Poor	4	2
	Average	4	7
	Best	5	4
Category 4: Provides professional factual-based feedback without applying judgement or emotional comments	Poor	0	0
	Average	1	0
	Best	12	13

Pre-service teachers' reflection analysis. Once the pre-service teachers completed the assessment of the two lab reports they were required to reflect on the assessment process and guidance and think what aspects of their 2nd assessment have changed compared to the 1st one. They had to respond to 7 reflection questions – 4 of which were related to using the rubric and the other 3 were related to providing written feedback to students. Each reflection was evaluated based on the categories shown in Table 5 and scored on a scale of Yes, Somewhat, and No / No comment. Under each category, the number of responses were summed by scores based on that scale. Majority of the pre-service teachers valued the assessment process as a learning process (category 2), as well as indicated that providing constructive feedback and scoring based on the rubric were

challenging tasks (categories 4 & 5). Especially challenging was evaluating and assessing the analysis, conclusions, and discussions dimensions (dimensions 3.2, 3.3, & 3.4) under section 3 of the lab report (category 6). However, just nearly half of them used the rubric differently and changed their feedback strategy for the 2nd assessment (categories 1 & 3).

*Table 5. Pre-service teachers' **reflection** analysis*

Assessment Categories of Pre-service teachers' Reflection	Yes	Some what	No / No comment
Category 1: I have given different constructive feedback to students across the 2 reports. *"…for the 2nd report, I knew better how to provide feedback and what to focus on…and be more objective. The 2nd time it was faster…"*	6	3	4
Category 2: Evaluating lab reports had been a learning process to me. *"Yes, it was easier for me to assess the 2nd report…I knew better what to expect…"*	11	1	1
Category 3: Have you used the Rubric differently across the 2 reports? *"Yes! The 2nd time I knew better what to look for…this is a skill I will have to develop…"*	6	5	2
Category 4: Providing constructive feedback is challenging. *"When something is missing/inaccurate, it is hard to just give an objective and constructive feedback…"*	9	4	0
Category 5: Scoring lab report based on the Rubric was challenging. *"Overall, it was challenging to score on a scale of 0 to 5. It is hard to choose…"*	11	1	1
Category 6: Evaluating the report's data analysis & conclusions dimensions under section 3 was challenging *"It is hard to evaluate...it is hard to decide what must be included vs. what is optional..."*	8	1	4

Contribution to teaching and learning science

Overall lab report score variation had been slightly lower because of the explicit guidance conducted by the professor. However, despite that guidance, variation had increased in results analysis, conclusions, and discussion dimensions, which were documented as complicated learning skills for students, and for teachers to assess (Berland & Reiser, 2009; McNeill & Krajcik, 2008). These dimensions are the epic of the lab report where students must clearly and concisely articulate their chemistry understanding inferred from an experiment and connect the macroscopic to the microscopic level (Bevins & Price, 2016). It is suggested that two major variables may cause the large variation: 1) Different assessors' mental understanding of what a complete explanation of chemistry understanding is (Bernard & Dudek-różycki, 2009), and 2) Scoring inconsistently due to rubric complexity (Allen & Tanner, 2006). Pre-service teachers' written feedback had improved because of the guidance across all categories, especially under the analysis, conclusions, and discussion dimensions category. This may suggest that assessors better articulated what must be improved, however scoring these dimensions remained an issue. Overall, the pre-service teachers valued the explicit guidance and it certainly made an impact in the way they approached the assessment of the 2nd report, however the least impact was on how they scored the 2nd report using the rubric. This research suggests that an emphasis should be put through research on developing high-school science teachers' assessment knowledge of inquiry-based laboratory reports (Author & colleagues, 2012), specifically in the conclusions and discussion dimensions.

Contribution to the interest of NARST members.

This study advances the idea that the use of rubrics should be considered as a potentially valuable assessment tool and feedback method. It sets out to examine the implementation and use of rubrics in the field of inquiry-based labs in chemistry education, as few papers examined this aspect in this discipline and in the context of high school and teachers' knowledge. We learned that rubrics provide explicit alignment to the aims of an assessment task. However, the research raises concerns of grading consistency even when using the rubric. Enhancing the formative assessment process may be achieved by creating a rubric which is simple and easy to use, as well as providing teachers with explicit guidance to promote their assessment knowledge.

Bibliography

Allen, D., & Tanner, K. (2006). Rubrics: Tools for Making Learning Goals and Evaluation Criteria Explicit for Both Teachers and Learners. *CBE*, *5*, 197–203. https://doi.org/10.1187/cbe.06 – 06 – 0168

Avargil, S., Bruce, M. R. M., Amar, F. G., & Bruce, A. E. (2015). Students' Understanding of Analogy after a CORE (Chemical Observations, Representations, Experimentation) Learning Cycle, General Chemistry Experiment. *Journal of Chemical Education*, *92*(10), 1626–1638. https://doi.org/10.1021/acs.jchemed.5b00230

Barneaa, N., Doria, Y. J., & Hofsteind, A. (2010). Development and implementation of inquiry-based and computerized-based laboratories: Reforming high school chemistry in Israel. *Chemistry Education Research and Practice*, *11*(3), 218–228. https://doi.org/10.1039/c005471m

Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, *93*(1), 26–55. https://doi.org/10.1002/sce.20286

Bernard, P., & Dudek-różycki, K. (2009). *Issn 1648-3898 Issn 2538-7138 Integration of Inquiry-Based Instruction With Formative Assessment : the Case of Experienced*. *April*, 184–196. https://doi.org/10.33225/jbse/19.18.184

Bevins, S., & Price, G. (2016). Reconceptualising inquiry in science education. *International Journal of Science Education*, *38*(1), 17–29. https://doi.org/10.1080/09500693.2015.1124300

Cacciatore, K. L., & Sevian, H. (2009). Incrementally approaching an inquiry lab curriculum: Can changing a single laboratory experiment improve student performance in general chemistry? *Journal of Chemical Education*, *86*(4), 498–505. https://doi.org/10.1021/ed086p498

Chiu, J. L., & Linn, M. C. (2014). Supporting Knowledge Integration in Chemistry with a Visualization-Enhanced Inquiry Unit. *Journal of Science Education and Technology*, *23*(1), 37–58. https://doi.org/10.1007/s10956-013-9449-5

Dori, Y. J., & Sasson, I. (2008). Chemical understanding and graphing skills in an honors case-based computerized chemistry laboratory environment: The value of bidirectional visual and textual representations. *Journal of Research in Science Teaching*, *45*(2), 219–250. https://doi.org/10.1002/tea.20197

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication & Technology*, *29*(2), 75–91. https://doi.org/10.1007/BF02766777

Hofstein, A., Dkeidek, I., Katchevitch, D., Nahum, T. L., Kipnis, M., Navon, O., Shore, R., Taitelbaum, D., & Mamlok-Naaman, R. (2019). Research on and Development of Inquiry-Type Chemistry Laboratories in Israel. *Israel Journal of Chemistry*, *59*(6), 514–523. https://doi.org/10.1002/ijch.201800056

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Huang, J., & He, Z. L. (2016). Exploring Assessment Literacy. *Higher Education of Social Science*, *11*(2), 18–27. https://doi.org/10.3968/8727

Kurdziel, J. P., Turner, J. A., Luft, J. A., & Roehrig, G. H. (2003). Graduate Teaching Assistants and Inquiry-Based Instruction: Implications for Graduate Teaching Assistant Training.

Journal of Chemical Education, *80*(10), 1206. https://doi.org/10.1021/ed080p1206
McNeill, K. L., & Krajcik, J. (2008). Inquiry and scientific explanations: Helping students use evidence and reasoning. *Science as Inquiry in the Secondary Setting*, 121–134.
Tashakkori, A., & Teddlie, C. (2015). Overview of Contemporary Issues in Mixed Methods Research. In A. M. Tashakkori & C. B. Teddlie (Eds.), *SAGE Handbook of Mixed Methods in Social & Behavioral Research* (pp. 1–42). SAGE Publications Inc. https://doi.org/10.4135/9781506335193.n1
van Brederode, M., Zoon, S., & Meeter, M. (2020). Examining the effect of lab instructions on students' critical thinking during a chemical inquiry practical. *Chemistry Education Research and Practice*. https://doi.org/10.1039/D0RP00020E
Walker, J. P., & Sampson, V. (2013). Learning to argue and arguing to learn: Argument-driven inquiry as a way to help undergraduate chemistry students learn how to construct arguments and engage in argumentation during a laboratory course. *Journal of Research in Science Teaching*, *50*(5), 561–596. https://doi.org/10.1002/tea.21082
Wheeler, L. B., Clark, C. P., & Grisham, C. M. (2017). Transforming a Traditional Laboratory to an Inquiry-Based Course: Importance of Training TAs when Redesigning a Curriculum. *Journal of Chemical Education*, *94*(8), 1019–1026. https://doi.org/10.1021/acs.jchemed.6b00831