

**Research Article**

Copyright © All rights are reserved by Chi Keung Chan

Developing Pre-service Chemistry Teachers' Assessment Literacy Using a Mentoring Scheme

Chi Keung Chan**Department of Science and Environmental Studies, The Education University of Hong Kong, Hong Kong*

***Corresponding author:** Chi Keung Chan, Department of Science and Environmental Studies, The Education University of Hong Kong, Hong Kong

Received Date: August 18, 2025**Published Date: August 26, 2025****Abstract**

Effective teaching in 21st-century education needs to infuse formative assessment tasks into students' personalized learning. This requires teachers to develop their assessment literacy so that they are able to design high-quality formative assessment tasks and provide students with constructive and timely feedback for meaningful conceptual change and self-directed learning. A project was conducted to develop pre-service science teachers' assessment literacy. The project consists of three stages: (1) The project team organized and conducted a series of workshops to provide participants with sufficient knowledge of the conceptual change, metacognition and application of multiple-choice questions (MCQs) for diagnostic purposes; (2) The participants were grouped and intensively shadowed by mentors to develop high-quality MCQs with feedbacks and produce well-planned short teaching videos; and (3) The participants were guided to make reflection in terms of assessor identity after studying the feedbacks from the secondary school teachers and students used the developed materials in their learning and teaching. To examine the project's effects, we developed a 20-item survey instrument that assesses participants' perceptions of their performance assessment literacy. The instrument covers three aspects: design of assessment tasks, providing students with feedback, and instruction based on assessment data. Pre-test and post-test were administered to study the project's effects, where participants rated their views on the items on a 5-point Likert scale. We found that participants' AL significantly developed after completing the project. Furthermore, the project produced a strong effect on the development of participants' AL concerning the design of assessment tasks.

Keywords: Assessment literacy; Diagnostic assessment; Alternative conception; Metacognition; Chemistry education

Introduction

Assessment literacy (AL) is regarded as a basic understanding of educational assessment and related skills to apply such knowledge to various measures of student achievement [1]. AL is now considered as integral to teacher professionalism [2] for two reasons. First, teachers are the key transforming agents in educational assessment [3]. Second, assessment is indivisible from student learning [4]. Consequently, teachers are required to make effective assessment-related decision-making, such as developing quality assessment items and providing feedback to facilitate

learning and teaching. In 21st-century classrooms, assessment tasks should provide students with an adaptive, personalized pathway for their self-directed learning [5].

Multiple-choice questions (MCQs) are commonly used as assessment items in public exams as well as daily learning and teaching in science education. There are two significant advantages of using MCQs in assessment tasks. First, we can assess learners' understanding of concepts fairly and efficiently if the MCQs are well-designed. Second, we can reveal learners' alternative conceptions/

misconceptions provided that the distractors are purposefully set [6]. It is very common for both students and teachers to use MCQs that appeared in past public exams during instruction and studies. However, although the MCQs from the past public exams are well-developed, they may be unable to facilitate effective learning and teaching for two significant reasons. First, MCQs in public exams are set for summative purposes, where the items are used to distinguish candidates into various levels and grades rather than facilitating learning and teaching. Second, some MCQs in the public exam are cross topics to assess candidates' knowledge in the mastering of knowledge of the whole curriculum after a few years of studies. Consequently, teachers are required to apply AL to adapt the MCQs from the public exam papers in daily learning and teaching. It would be more meaningful if teachers could exercise their AL to develop MCQs to meet student learning needs.

An MCQ can be an effective learning tool, especially when it is able to induce a learner's conceptual change [7-9] and metacognition [10-12]. An MCQ is able to achieve this goal when the following two features are included. First, the distractors are set based on learners' possible alternative conceptions/misconceptions. Second, immediate and constructive feedback is provided to learners when they are not able to select the correct answer. Additionally, some educationists [13-15] advocate the use of two-tier MCQs as diagnostic instruments where learners' factual knowledge and conceptual knowledge are to be assessed in the first and second tier respectively. In short, a high-quality MCQ is able to facilitate effective learning, teaching and assessment. Yet, significant effort is required to develop high-quality MCQs because it involves practical application of AL [16].

However, key stakeholders such as school principals, experienced teachers and teacher educators constantly criticized inexperienced and pre-service teachers who performed relatively weakly in designing assessment tasks and providing feedback. In this connection, nurturing pre-service teachers' AL in teacher education programmes is essential. As a result, we referred to the conceptual framework of Teacher Assessment Literacy in Practice (TALiP) [16], which consists of three levels of mastery when designing and implementing our teacher development project.

Methodology

Participants and Procedures

The participants were 31 student-teachers. Of the 31 of them, 20 (65%) were male and 11 (35%) were female. Twenty-eight (90%) of them majored in science education, specializing in chemistry in their five-year full-time undergraduate degree, while the remaining 3 (10%) majored in science education, specializing in chemistry in their two-year part-time postgraduate diploma of education. The former were prospective teachers, while the latter were in-service chemistry teachers with less than 4 years of teaching experience.

The project team comprised a teacher educator (the author), a former curriculum development officer, a former senior manager of assessment development, a retired secondary school principal with chemistry teaching experience, and two in-service teachers with

more than 20 years of experience in teaching chemistry.

At the beginning of the teacher development project, we provided each participant with a printed survey questionnaire (hard copy) to complete. After finishing the project, the participants took the survey again. The two sets of surveys were identical and can be regarded as pre- and post-tests on their perceptions of their AL.

The Teacher Development Project

I am the lecturer teaching the chemistry pedagogy course at the undergraduate and postgraduate levels. Students were invited to join the teacher development project on a volunteer basis. The project consists of two phases, which can be further divided into three stages.

In Stage 1 of each phase, the project team organized and conducted a briefing session and a series of workshops to provide participants with sufficient knowledge of the conceptual change, metacognition and application of MCQs for diagnostic assessment. The total duration of the briefing session and workshops was 9 hours. In this stage, participants joined the project as the participants.

In Stage 2 of each phase, the project team members grouped and intensively shadowed participants to develop high-quality MCQs with written feedback and produce well-planned short teaching videos. The total duration of this mentoring stage was 6 months. In this stage, participants engaged in the project as the co-developers.

In Stage 3 of each phase, the project team developed an open online platform to contain all the developed MCQs and materials. Chemistry teachers and students teaching and studying the HKDSE Chemistry (equivalent to grades 10-12) in two secondary schools used the platform. The participants were guided to revise the materials according to the feedback gathered from the secondary schools. Furthermore, a moderation meeting and a debriefing session were held to facilitate them to make reflections on the development of their AL. The total duration of the meeting and debriefing session was 5 hours. In this stage, participants engaged in the project as reflective practitioners.

We suffered from the constraint of a small class size and thus divided the project into two identical phases. A total of 31 students from the undergraduate and postgraduate diploma of education programmes participated in the project, where 19 participated in phase 1 and 12 participated in phase 2 (for details of the project, please see [17]). Through their active and collaborative engagement in the project, they developed their AL in practice, in particular:

- (1) They developed 208 valid MCQs regarding specific learning objectives;
- (2) They developed written feedback for each distractor to induce the learner's conceptual change and metacognition;
- (3) They developed 20 short teaching videos to explain the concepts involved in the selected MCQs; and

(4) We developed an open online platform that included all the MCQs, feedback, and videos for secondary school teachers and students.

The Survey Instrument

To study the impacts of the teacher development project on participants' AL, we developed a survey instrument composed of 20 items to cover three aspects: (1) design of assessment tasks, (2) providing students with feedback, and (3) instruction based on assessment data.

The first principle guiding the design and development of the MCQs is that the items can facilitate students' learning. Furthermore, the MCQs should validly align with the assessment objectives. The second principle is that the MCQs should provide students with written feedback to induce conceptual change and metacognition. Finally, the third principle is that teachers refer to students' performance in the MCQs to design or adjust their teaching.

We, based on the three principles and referred to the PAL scale [18], developed a 20-item survey instrument for our project. The items are presented in Table 1. Meanwhile, the items were randomly distributed in the survey questionnaire. We asked the participants

to rate their degree of agreeability on each item on a five-point scale from 1 (very disagree) to 5 (very agree).

Item analysis was conducted on the 20 items. The Cronbach's alphas of the items in the pre-test and post-test were 0.87 and 0.82, respectively. The values were greater than 0.7, suggesting that internal consistency was obtained among the items [19].

Data Analysis

We investigated whether the participants significantly improved in AL after the teacher development project. A paired-samples t-test was conducted to analyse the effectiveness of the teacher development project. Levene's tests of equality of error variances were conducted to check whether the statistical assumption of homogeneous variance was violated. The above analysis was conducted using SPSS 29.0. In case there was a violation of statistical assumption, we conducted the Wilcoxon test, which is a non-parametric test to analyse the main and interaction effects for the data with the violation of statistical assumptions [20]. The effect size of a t-test was indicated using Cohen's d, with a range of 0-0.20 indicating a weak effect, 0.21-0.50 a modest effect, 0.51-1.00 a moderate effect and equal to and greater than 1.01 a strong effect [19].

Table 1: Survey instrument items.

Item id	Question no in the survey	Item text
D_1	1	I am able to design assessment tasks as diagnostic instrument of student learning
D_2	4	I am able to design assessment tasks that provide students with feedback to make decisions about their learning
D_3	5	I am confident to develop assessment tasks that are clearly aligned to the curriculum
D_4	8	I am able to design assessment tasks that allow students to monitor their own learning
D_5	9	I am able to develop assessment tasks to assess students' understanding of knowledge
D_6	12	I am able to develop assessment tasks that provide students with opportunities to reflect on their learning
D_7	13	I am able to design assessment tasks to identify student's alternative conceptions/misconceptions
D_8	16	I am able to design assessment tasks that promote an academic growth mindset (i.e., the belief that their academic abilities are not fixed and can improve with effort)
D_9	7	I am confident to develop assessment tasks that are clearly aligned with state objectives
D_10	20	I am confident to design assessment tasks to promote student learning
F_1	2	I am able to provide students with concrete written feedback to facilitate their learning
F_2	6	I am able to write specific feedback based on students' performance in assessment tasks
F_3	10	I am confident to provide students with feedback to facilitate their conceptual change
F_4	14	I am confident to provide students with feedback to facilitate their metacognition
F_5	18	I am able to write feedback to facilitate students' learning
I_1	3	I am able to modify instruction for students based on student assessment data
I_2	7	I am able to adjust instruction for particular groups of students based on student assessment data
I_3	11	I am able to provide individual students with personalized instruction based on student assessment data
I_4	15	I am able to refer to student assessment data to design/plan my lessons
I_5	19	I am confident to develop teaching videos to explain the concepts that are related to the assessment tasks

Results

Overall Analysis

We conducted a paired-samples t-test to evaluate the effectiveness of the teacher development project on participants' AL. In general, the results indicated that the mean of post-test ($M = 79.06$, $SD = 6.23$) was significantly greater than the mean of pre-test ($M = 68.83$, $SD = 8.79$), $t(17) = 3.74$, $p < .01$. The standardized effect size index, d , was .88, with very little overlap in the distributions for the 5-point Likert ratings in the survey instrument, as shown in Figure 1. The 95% confidence interval for the mean difference between the two ratings was 4.46 to 15.99.

The results supported that the participants self-perceived their AL, which had been significantly developed after completing the teacher development project.

Aspects Analysis

The 20 items in the survey instrument covered three aspects: (1) design of assessment tasks, (2) providing students with feedback, and (3) instruction based on assessment data. We conducted a paired-samples t-test to analyse which aspects contributed to the large effect of the teacher development project, as reported in Section 3.1.

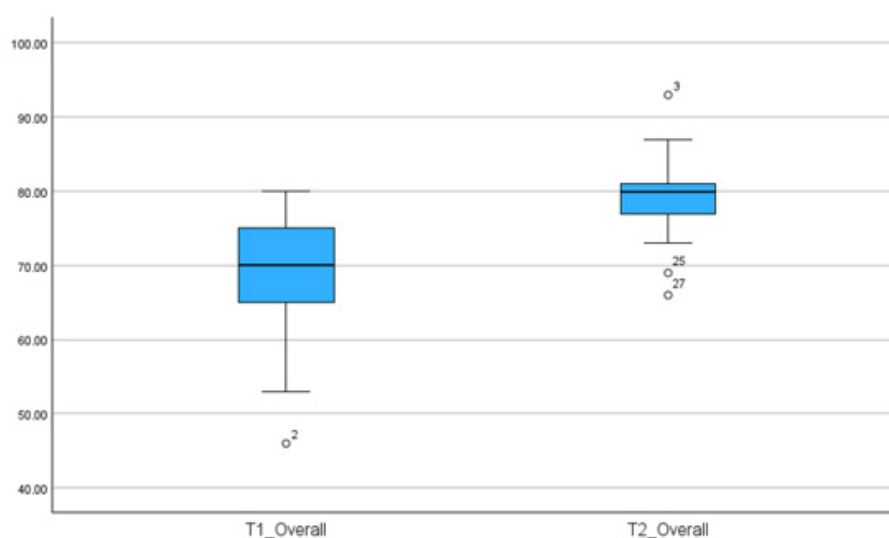


Figure 1: Boxplots of pre-test (T1) and post-test (T2) ratings.

Table 2: Results of the paired-samples t test of the pre-test (T1) and post-test (T2) by the three aspects and overall ratings.

Pair	Mean	SD	95% Confidence interval of the difference		t	df	Significance		d
			Lower	Upper			One-sided p	Two-sided p	
T2_Design – T1_Design	6.28	6.22	3.18	9.37	4.28	17	<.001	<.001	1.01
T2_Feedback – T1_Feedback	1.67	3.53	-0.09	3.42	2	17	0.031	0.061	0.47
T2_Instruction – T1_Instruction	2.28	3.69	0.44	4.11	2.62	17	0.009	0.018	0.62
T2_Overall – T1_Overall	10.22	11.59	4.46	15.99	3.74	17	<.001	0.002	0.88

Results of the paired-samples t-test of the pre-test and post-test by the three aspects and overall ratings were presented in Table 2.

The results indicated that the teacher development project produced a strong effect on the development of participants' AL in terms of the design of assessment tasks, a moderate effect in terms of instruction based on assessment data, and nearly a moderate effect in terms of providing students with feedback.

Conclusion

Possible Reasons for the Results

The results are encouraging and rewarding to the project participants and team members. The three-staged project generally provided participants with holistic and comprehensive learning experiences for active engagement.

Second, participants were intensively shadowed by the team members in developing high-quality MCQs and feedback in stage 2 of the project. Participants' active engagement in the learning process may be the reason to explain why they rated the design aspect significantly higher in T2 than in T1.

Meanwhile, modifying instruction and providing students with feedback based on assessment data may require an authentic teaching context. On the other hand, developing MCQs and related materials may not fully develop these two aspects. Consequently, participants rated their AL in these two aspects with a medium effect.

Implications of the Project

We analysed the effectiveness of a teacher development project on developing pre-service science teachers' AL. We found that the participating teachers' AL were developed with moderate to strong effects. The findings suggest that in designing similar teacher development programmes of AL, it is essential to nurture their competencies to design valid assessment tasks to facilitate students' learning and align with the objectives. Furthermore, developing teachers' AL requires developing their competencies to provide constructive feedback for student's conceptual change and metacognition. Additionally, it is essential to develop teachers' AL regarding using assessment data to enhance teaching. Findings from our project further support the idea that assessment is indivisible from student learning and teacher instruction.

Limitations of this Study

This study has three significant limitations. First, all the participants were recruited and joined this teacher development project on a volunteer basis. It is very likely that all the participants were highly motivated and had confidence in designing MCQs and writing feedback for chemistry learning and teaching. Consequently, they self-perceived their AL quite well, which created a generalizability problem.

Second, there were 31 participants joined the project. We were able to collect 30 questionnaires from the participants in the pre-test because one participant was absent from the briefing session. However, we only received 18 questionnaires from the participants in the post-test because 13 participants did not return the questionnaires after they had graduated from their studies. Although we have used statistical tools to ensure the assumption of homogeneous variance was not violated, the small sample size may create a generalizability problem.

Finally, we did not employ an experimental design using a control group. We noted this limitation at the beginning. However, the small class size of the chemistry pedagogy course involved limited the recruitment of participants. Indeed, almost all the students in the course had joined the project. Then, it might not be appropriate to assign half of the participants. The reason is that there would be a reduction in sample size if the current group of participants (N=31) were split into an experimental and a control group. A possible modification in future studies is to recruit some teachers as the control to explore whether there are confounding

variables (e.g., accumulation of teaching experiences) on the development of AL independent of the teachers' intervention [21-28].

Acknowledgements

The author acknowledges that the funding for this Teaching Development Grants (TDG) project was supported by the Education University of Hong Kong (EdUHK).

Conflict of Interest

No conflict of interest.

References

1. Stiggins RJ (1991) Assessment literacy. *Phi Delta Kappan* 72: 534-539.
2. Engelsen KS, Smith K (2014) Assessment literacy. C Wyatt-Smith, V Klenowski, P Colbert (Eds.), *The Enabling Power of Assessment: Designing Assessment for Quality Learning*. New York: Springer pp. 140-162.
3. OECD (2019) *OECD Future of Education and Skills 2030 Concept Note*.
4. Black P, William D (1998) Assessment and classroom learning. *Assessment in Education* 5(1): 7-74.
5. Zimmerman BJ, Martinez-Pons M (1990) Student differences in self-regulated learning: relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology* 82(1): 51-59.
6. Treagust DF (1988) Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education* 10: 159-169.
7. Carey S (1991) Knowledge acquisition: Enrichment or conceptual change? In S Carey, R Gelman (Eds.), *The Epigenesis of Mind*. Hillsdale, NJ: Erlbaum pp. 257-291.
8. Chi MTH (1992) Conceptual change within and across ontological categories: Examples from learning and discovery in science. In RN Giere (Ed.), *Minnesota Studies in the Philosophy of Science*, Vol. XV: Cognitive Models of Science. Minneapolis: University of Minnesota Press pp. 129-186.
9. Chi MTH (2008) Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In Vosniadou, S. (Ed.), *International Handbook of Research on Conceptual Change*. New York: Routledge pp. 61-82.
10. Brown A (1987) Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In FE Weinert, RH Kluwe (Eds.), *Metacognition, Motivation, and Understanding*. Hillsdale, New Jersey: Lawrence Erlbaum Associates pp. 65-116.
11. Flavell JH (1987) Speculations about the nature and development of metacognition. FE Weinert, RH Kluwe (Eds.), *Metacognition, Motivation, and Understanding* (21-29). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
12. Kuhn D, Dean D (2004) Metacognition: A bridge between cognitive psychology and educational practice. *Theory into Practice* 43(4): 268-276.
13. Chou CC, Chiu MH (2004) A two-tier diagnostic instrument on the molecular representations of chemistry: Comparison of performance between junior high school and senior high school students in Taiwan. Paper presented at the 18th International Conference on Chemical Education, Istanbul, Turkey.
14. Chandrasegaran AL, Treagust DF, Mocerino M (2007) The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice* 8(3): 293-307.

15. Adadan E, Savasci F (2012) An analysis of 16-17-year-old students' understanding of solution chemistry concepts using a two-tier diagnostic instrument. *International Journal of Science Education* 34(4): 513-544.
16. Xu Y, Brown GTL (2016) Teacher assessment literacy in practice: A reconceptualization. *Teacher and Teacher Education* 58: 149-162.
17. Chan CK (2024) Developing assessment literacy of pre-service science teachers in practice. In *Proceedings of the 17th International Conference of Education, Research and Innovation (ICERI2024)*. Seville, Spain. IATED.
18. Kelly MP, Feistman R, Dodge E, St Rose A, Littenberg-Tobias J (2020) Exploring the dimensionality of self-perceived performance assessment literacy (PAL). *Educational Assessment, Evaluation and Accountability* 32: 499-517.
19. Muijs D (2022) *Doing Quantitative Research in Education with IBM SPSS Statistics* (3rd ed.). SAGE.
20. Green SB, Salkind NJ (2014) *Using SPSS for Windows and Macintosh: Analysis and Understanding Data*. Upper Saddle River, New Jersey: Pearson.
21. Curriculum Development Council (2017) *Secondary Education Curriculum Guide*. Hong Kong: Government Logistics Department.
22. Curriculum Development Council & Hong Kong Examinations and Assessment Authority (2007) *Chemistry Curriculum and Assessment Guide (Secondary 4-6)*. Hong Kong: Government Logistics Department.
23. Curriculum Development Council & Hong Kong Examinations and Assessment Authority (2015) *Chemistry Curriculum and Assessment Guide (Secondary 4-6)*. Hong Kong: Government Logistics Department.
24. Harrison GM, Vallin LM (2018) Evaluating the metacognitive awareness inventory using empirical factor-structure evidence. *Metacognition and Learning* 13(1): 15-38.
25. Liao YW, She HC (2009) Enhancing eight grade students' scientific conceptual change and scientific reasoning through a web-based learning program. *Educational Technology & Society* 12(4): 228-240.
26. Shulman LS (1987) Knowledge and training: Foundations of the new reform. *Harvard Educational Review* 57(1): 1-22.
27. Wang TH (2014) Developing an assessment-centred e-Learning system for improving student learning effectiveness. *Computers & Education* 73: 189-203.
28. Yuruk N (2005) An analysis of the nature of students' metaconceptual processes and the effectiveness of metaconceptual teaching practices on students' conceptual understanding of force and motion. *Dissertation Abstracts International* 66(07): 2485A.