

Received: 25 May 2024, Accepted: 28 July 2024

**From Observation to Instructional Change: Patterns of Teaching Practice
Across Five Feedback Cycles in Grade VIII Science and Mathematics
Classrooms**

Kazma Saleem

MPhil Scholar Institute of Education and Research University of the Punjab, Lahore, Pakistan.

kazmasaleem52@gmail.com

Abstract

This study reports on the variation of four observed instructional practices over five cycles of classroom observation and feedback. The Teach observation sheet recorded lesson facilitation, checks for understanding, instructional feedback and critical thinking at five observation points. Analysis was conducted to report on the dissertation's descriptive statistics and selected school-level tests without re-computing results from unavailable individual tests. All four mean scores were higher at observation five than at observation one at the full-sample level. Lesson facilitation rose from 12.17 to 15.10, checks for understanding from 10.28 to 12.00, instructional feedback from 6.96 to 7.64 and critical thinking from 10.57 to 12.28. The paths differed. There was a fall in lesson facilitation and instructional feedback at the second observation, though it did improve. Feedback and critical thinking also decreased slightly between observations 4 and 5. The thesis reported an omnibus result for lesson facilitation for School 3 as $F = 5.77$, $p = .05$. No school-level omnibus p value was less than .05 for the other instructional elements. These are describing inequitable changes in observed practice. The single group design cannot prove that the changes were due to observer feedback. The paper provides a focused narrative about the timing and stability of instructional change, and areas that might need continued coaching.

Keywords: Classroom Observation, Observer Feedback, Instructional Practice. Grade VIII, Science Teaching, Mathematics Teaching, Repeated Measurement

1. Introduction

Teachers can observe during a class to see the manner in which teaching takes place. A structured instrument renders practices visible. Feedback following observation can then be used to guide attention to a practical next step. However, a rise from a first to a final score only provides part of the picture. The strengthening process of a practice can be smooth and continuous or can be

followed by a dip before it strengthens or can be accompanied by an increase before a weakening. These patterns are important when the school makes decisions about the frequency of lesson observations and the type of support that teachers require. Observations can identify the point in a series of observations at which a practice begins to change and whether the change is still discernable in subsequent lessons. Science and mathematics teachers deal with concepts which a pupil may only partially understand.

Teachers must be able to clearly explain ideas, explore pupils' understanding, listen to their responses, and allow time for reasoning. Although these actions are connected, they are not the same. A teacher can plan a lesson in a more effective manner and still struggle to engage pupils' thinking and respond to it. This difference may be masked by a single overall effectiveness score. The evidence on feedback also highlights the need for useful information to include information relating to current performance and the next step (Hattie & Timperley, 2007). This means that classroom observations should make a distinction, therefore, between practices that are being improved and practices that require more support. The dissertation which generated this article analyzed teachers during five classroom observations. Changes in nine elements of classroom culture, instruction and socioemotional skills were summarised in its earlier article (Saleem & Abiodullah, article provided by the authors). The following analysis addresses a more limited question: What were the changes in the four observed features of the teaching/learning as observed from one time to another? It focuses on lesson facilitation, checking for understanding, instructional feedback and critical thinking. The objective is to find out when change was recorded, and where and how stable it was. This dissertation is the original source for all data. These are a focused secondary analysis of one sample and not an independent replication or intervention.

2. Literature Review and Conceptual Basis

Feedback can be used to facilitate professional learning if it is specific enough to elicit action from the teacher. Feedback is information about progress towards a goal and the next steps that should be taken (Hattie and Timperley, 2007). This involves making a connection between a comment and a visible aspect of teaching in one observation cycle. For example, a comment on questioning will be more helpful if it outlines the questions posed and the opportunities for pupils to present their reasons. If teachers are expected to adjust classroom practice, their comments on observations should be focused. A classroom observation score is most useful when followed up by a discussion of evidence. Teachers should be informed about what action they saw, why it is important for learning and what they might consider doing next. The same goes for feedback from an observer to

a teacher and from a teacher to pupils. The dissertation quantifies the latter as a component of observed teaching. It documents the former in the documentation of the observation but does not quantify the content. Separating the two meanings avoids an unsupported claim that more observer feedback necessarily led to better feedback to pupils (Hattie & Timperley, 2007).

Multiple professional learning offers a chance to revisit a practice following an attempt. In the dissertation, there were several observation points for the teachers and not a final judgment. This will allow changes to be shown in the intermediate structure. It also aligns with research that views mentoring and feedback as a process (Keiler et al., 2020). However, we cannot determine from a series of scores whether a teacher used a specific recommendation. The rating can also be subject to change, due to changes within the topic, in the activity of the class, or in the judgment of the observer. The importance of repeated design is not to exclude all other possibilities but to make it clear what the pattern is.

Instructional elements may vary in terms of how quickly they change, as they have varying demands on teachers. Facilitation of lessons is apparent in the organisation and pacing of activities. Checks for understanding involve finding out what pupils are thinking now. Feedback to pupils must be responded to in instructional feedback. Learning opportunities that call for explanation and justification are essential in critical thinking. Teacher noticing research indicates close attention to pupils' ideas (van Es & Sherin, 2008) while critical thinking research suggests that critical thinking needs deliberate teaching strategies (Abrami et al., 2015). These are sources of interpretation and do not contain any more data for the present sample.

There are limits to observation protocols, too. A rating may indicate the presence or quality of a classroom practice within the lesson segments observed, but not all aspects of instruction. More checks for understanding will not automatically indicate more understanding of the lesson. Similarly, the higher the critical thinking score the less it directly measures pupil reasoning. Literature on observation protocols highlights the need to understand what an instrument observes and the domains that make up the instrument (Anwar & Menekse, 2021). This paper therefore has kept the four element labels from the dissertation and now views these as observable teaching practices.

3. Method

3.1 Design and Setting

This secondary analysis uses the dissertation's single-group, repeated-observation design. The context was four public schools in Lahore, Pakistan. There were five observations of teachers with feedback from observations between the five. The design described in the thesis is a quasi-experimental design. It does not have a comparison group, and results are better thought of as changes documented as part of an observation and feedback process. The grade is given here as Grade VIII as confirmed by the researcher, several parts in the dissertation text are still indicated as Grade VI and needs to be corrected against the original study records before submission.

3.2 Participants and Sampling

The analytic sample of the first article consists of 28 teachers: eight teachers from School 1, eight teachers from School 2, eight teachers from School 3 and four teachers from School 4. The dissertation is based on a convenience school sample and only covers the role of science and mathematics teaching. There is no distinction made between science and mathematics since the summary tables provided do not reflect a subject-specific distinction that has been verified. Scores thus apply to the total sample.

3.3 Observation Instrument

The observation sheet from the Teach Observer Manual was used for the dissertation. There are 28 statements across 9 elements and 3 overall domains of classroom culture, instruction and socioemotional skills. This article employs only the four aspects of instruction. The researcher observed two 15-minute segments of the lesson each time and coded the observed practice on the instrument's 5-point rating system. The scores in this paper represent the sum of figures for items on a figure, not the individual item ratings or percentage scores of the dissertation. A comparison of the raw size of different element scores is inappropriate, if all four scales had the same range.

3.4 Procedure

The researcher gained access to the school and observed lessons and filled out the checklist. FBCs were conducted following observation with a short feedback conversation. Then the same teaching categories were noted at subsequent observation points. A brief inter-round interval is mentioned in the dissertation and five observation points are described. It won't keep a record of all the comments

that are made to each teacher. Thus, this present paper cannot relate a particular comment to a particular subsequent score. It investigates the order of the recorded scores of instructions instead.

3.5 Statistical Analysis

The four instructional elements are exactly represented in the dissertation's Table 4.2 means and standard deviations. It covers the changes from one consecutive observation to the next and compares the first and fifth observation in a descriptive manner. When the interpretation of selected school-level F statistics and paired t-test results provides information to explain a pattern, the results are reported. An inferential test, effect size, confidence interval or participant-level model is not developed because the individual teacher records were not provided. The difference between two reported means is not a separately tested difference, but rather is a descriptive arithmetic difference. P-values are reported for several pairwise tests, and values may be inflated due to multiple testing, so caution should be used when interpreting them individually.

3.6 Ethical Considerations

The dissertation states that the researcher obtained permission from school principals and relevant teachers, informed teachers about participation and kept their data confidential. This is a secondary article and only summarizes groups and school numbers, not individual identities. It is only based on the findings of the dissertation. There was no new contact with participants, and no new data was collected.

4. Results

The full sample instruction data, as presented in the dissertation's Table 4.2, is provided in Table 1. Mean and standard deviation (SD) are reported for each cell. All four final means > their initial means. However, the time periods between observations vary by elements. Table 1. Observation 1-5: M (SD) Instructional element scores

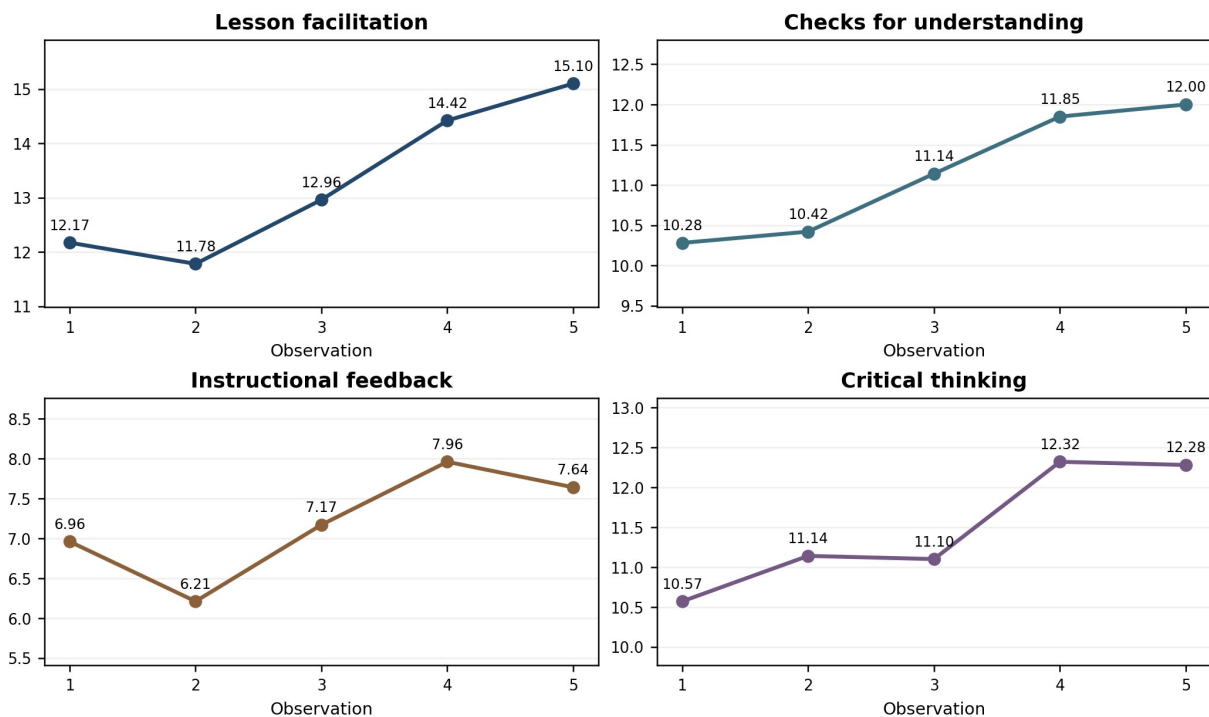
Table 1. Instructional element scores across five observations, M (SD)

Instructional element	Obs. 1	Obs. 2	Obs. 3	Obs. 4	Obs. 5
Lesson facilitation	12.17 (3.65)	11.78 (3.78)	12.96 (5.00)	14.42 (3.87)	15.10 (3.75)
Checks for	10.28 (2.78)	10.42 (3.29)	11.14 (2.95)	11.85 (2.25)	12.00 (1.58)

understanding					
Instructional feedback	6.96 (2.26)	6.21 (2.52)	7.17 (2.73)	7.96 (1.75)	7.64 (1.87)
Critical thinking	10.57 (3.72)	11.14 (3.00)	11.10 (3.47)	12.32 (2.14)	12.28 (2.10)

Note. Values are transcribed from the dissertation's instruction summary table (Table 4.2). "Feedback" in that table refers to teacher feedback to pupils as scored during classroom observation; it is distinct from feedback given by the observer to the teacher.

Figure 1. Mean instruction scores across five observation points. Each panel uses its own y-axis scale; compare changes within an element, not the height of different elements.



4.1 Lesson Facilitation

Lesson facilitation increased at 12.17 (SD = 3.65) and decreased at the second observation to 11.78 (SD = 3.78). It then increased to 12.96 (SD = 5.00), 14.42 (SD = 3.87) and 15.10 (SD = 3.75) at observations three, four and five. The mean difference between the first to fifth score was 2.93 points. So, there was first a decrease and then the increase. This movement before the endpoints would otherwise not be evident from a report based just on the endpoints. There is also significant variation in scores at this point as shown by the high third-observation standard deviation, but this does not reveal which teachers may be the cause of that variation. The result from the thesis is an omnibus school-level result: $F = 5.77$, $p = .05$ (School 3). Its means for that school were 10.00,

11.62, 14.00, 16.12 and 16.50. The omnibus p values for Schools 1, 2 and 4 were .42, .94, and .42, respectively. Between the first and last observation, School 1 rose from 10.00 to 14.00, School 2 dropped from 14.12 to 13.87, and School 4 remained at 17.00. These various endpoints provide some context as to why a pooled increase should not be interpreted as an increase in each school. In the thesis, the paired comparison of School 3, 1st versus 5th, is given as $t = -5.30$, $p = .00$; the manuscript should read $p < .01$ and not 0. The paper considers that comparison exploratory, as the source does not give unrounded p values or make any adjustment for multiple testing.

4.2 Checks for Understanding

The most orderly full-sample sequence was that of checks for understanding. The means were 10.28, 10.42, 11.14, 11.85 and 12.00. All observations were 10% or more above previous observations. The difference between the first and fifth descriptive was 1.72 points. The standard deviation has gone down from 2.78 at observation one to 1.58 at observation five. This is due to the higher mean and smaller final spread, indicating that scores were more concentrated by the final observation. The table does not reflect the actual attainment of pupils, only what was observed about teacher's practice and quality of questioning.

4.3 Instructional Feedback

Instructional feedback to pupils began at 6.96 (SD = 2.26), fell to 6.21 (SD = 2.52), rose to 7.17 (SD = 2.73) and 7.96 (SD = 1.75), then ended at 7.64 (SD = 1.87). An endpoint difference of 0.68 score points was found. This was the smallest raw difference between the first and the fifth elements; but, because the elements are measured on different scales, it is not valid to make a ranking of the standardized improvement based on the raw difference. Final score was above baseline, but below 4th observation. The sequence indicates a partial conclusion; the feedback practice observed was not consistent across the sequence and the peak of the fourth round was not carried over fully into the final round.

4.4 Critical Thinking

Critical thinking scores were 10.57 (SD = 3.72), 11.14 (SD = 3.00), 11.10 (SD = 3.47), 12.32 (SD = 2.14) and 12.28 (SD = 2.10). The difference between the first and fifth was 1.71 points. There were slight decreases between observations two and three and between four and five. A higher score on the observation instrument could have reflected more opportunities to support reasoning observed, but the study has not tested whether pupils' own critical thinking has improved. It is important in

science and mathematics where classroom opportunities and learner outcomes are separately measured concepts.

The thesis presents the school-level omnibus p values: .11, .92, .35 and .92. It also includes mismatched paired-test data for this element. For instance, the fourth and fifth descriptive values are the same in School 1 but the paired-test table for that pair does not have a t value of 0; another pair of first and fifth descriptive values has a positive t value but an increasing mean. Identical rounded group does not exclude individual differences at the individual level but cannot be checked without the individual scores or the original statistical output. This article therefore does not interpret the paired tests as secure evidence and is based on the full-sample descriptive table.

4.5 Summary of School Level Omnibus Results

The omnibus tests for each of the four elements of instruction at the school level are reported in the thesis and only the test for lesson facilitation reaches the thesis reported boundary at a $p = .05$ level. All school-level omnibus p values for checks for understanding, instructional feedback, and critical thinking are not significant. The findings demonstrate a helpful separation of direction and statistical evidence. A higher final mean describes what is recorded, and variation amongst observations under the statistical assumptions of an omnibus test is examined. The source does not provide sufficient detail to audit all the assumptions, specifications or corrections for multiple comparisons. The paper thus stresses the sequences observed and views the statistical results with caution.

Table 2. School-level omnibus statistics reported in the dissertation

Instructional element	School 1 F / p	School 2 F / p	School 3 F / p	School 4 F / p
Lesson facilitation	1.23 / .42	.16 / .94	5.77 / .05	.78 / .42
Checks for understanding	1.25 / .41	1.72 / .30	1.56 / .33	1.45 / .44
Instructional feedback	1.73 / .30	.47 / .75	1.67 / .31	.47 / .75
Critical thinking	3.68 / .11	.21 / .92	1.49 / .35	.21 / .92

5. Discussion

This purposeful analysis reveals that no single pathway for instructional practice emerged. Means of all four elements increased at the 5th observation as compared to the 1st observation. However,

lesson facilitation and instructional feedback were at the second observation. There was almost a level trend in critical thinking between the second and third observations and a slight decrease in the final observation. Checks for understanding alone increased at each complete observation of the sample. Such differences are significant when a first-to-last comparison can make a complex learning process appear smooth. Instructional feedback was not as consistent.

The second was lower than the baseline, and the last one was lower than the fourth-round high. Observer feedback given to teachers and instructional feedback given by teachers to pupils are also different activities. This paper documents the latter as an outcome and the former as a study procedure. It is unable to find a line of sight from one kind of feedback to its other kind. The distinction serves to enhance the interpretation because it means that the score change is not ascribable to a mechanism not directly assessed in the dissertation.

5.1 Strengths and Limitations

There are several limits to inference. The convenience sampling and the number of teachers (28) restricts the generalization. There is no untreated comparison for the single group design, and some of the change may be due to the passage of time, lesson content, familiarity with the instrument and other school events. The available tables do not include teacher level information, observer comments, content information per subject, student achievement or inter-rater agreement. According to the thesis, reliability was determined, but the methodological passage checked for this article does not specify a coefficient that can be verified here. Scores therefore should be regarded as observations made by the scorer, and not as exact measurements free from scorer error. There are also inconsistencies in the dissertation which require being resolved prior to journal submission. It mentions

Grade VI several times in its prose; the researcher notes Grade VIII. Different values are presented for some of the observations of the instructions in the first article compared to the main Table 4.2 in the dissertation. There are some internally inconsistent paired test entries, and some p values reported in the thesis are .00. This paper retains the information presented in Table 4.2, cites the source of each statistic presented and does not make claims from questionable statistics. The authors should compare the manuscript with original SPSS output and participant records before submitting it.

6. Conclusion

All four full-sample scores for instruction were above their baseline value after 5 observation and feedback cycles. They had divergent lives. There were at least one interim decline in lesson facilitation, instructional feedback and critical thinking, and checks for understanding increased at each observation. The thesis' clearest reported omnibus result for lesson facilitation was from School 3 ($F = 5.77$, $p = .05$). The implications of these findings are for a practical interpretation of repeated observation, where the timing and stability of change is as important as the end result. They are not evidence of feedback leading to improvement, or pupils learning more. Future research should include retention of feedback records and the linking of observed instructional practices with direct measures of learning.

References

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research*, 85(2), 275–314.
- Anwar, S., & Menekse, M. (2021). A systematic review of observation protocols used in postsecondary STEM classrooms. *Review of Education*, 9(1), 81–120.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915–945.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Keiler, L. S., Diotti, R., Hudon, K., & Ransom, J. C. (2020). The role of feedback in teacher mentoring: How coaches, peers, and students affect teacher change. *Mentoring & Tutoring: Partnership in Learning*, 28(2), 126–155.
- Saleem, K. (2025). Examining the influence of observer feedback on effectiveness of teaching science and mathematics for Grade VI (MPhil thesis, University of the Punjab). Unpublished thesis.
- Saleem, K., & Abiodullah, M. (n.d.). Observer feedback and changes in teaching effectiveness among Grade VIII science and mathematics teachers: A quasi experimental study in Lahore, Pakistan.
- van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers' learning to notice in the context of a video club. *Teaching and Teacher Education*, 24(2), 244–276.