

Observer Feedback and Changes in Teaching Effectiveness among Grade VIII Science and Mathematics Teachers: A Quasi Experimental Study in Lahore, Pakistan

Kazma Saleem¹, Dr. Muhammad Abiodullah²

1. MPhil Scholar Institute of Education and Research University of the Punjab, Lahore, Pakistan. E-mail: kazmasaleem52@gmail.com
2. Associate Professor Institute of Education and Research, University of the Punjab, Lahore, Pakistan. E-mail: abiodullah.ier@pu.edu.pk

Corresponding author: E-mail: kazmasaleem52@gmail.com

Abstract

The present study explored the level of observed teaching effectiveness in Grade VIII science and mathematics teacher of four public schools of Lahore, Pakistan through multiple cycles of observation and structured feedback by an observer. A single group repeated observation quasi experimental design was used for the study. The number of teachers included in the school level statistical tables is an analytic sample of 28 teachers, consisting of 8 teachers each in School 1, School 2, School 3 and 4 teachers in School 4. Teaching has been evaluated from five Observation Sheet points in the Teach Observer Manual (2nd edition). Three domains and nine elements were included in the instrument: classroom culture, instruction, and socioemotional skills. Mean and standard deviations were calculated at each observation point and paired sample t tests and analysis of variance were used for comparisons of the school levels. For each of the nine elements, the mean scores at Observation 5 were higher than mean scores at Observation 1, although some elements were not monotonic. There were clear descriptive gains in lesson facilitation, perseverance, and social and collaborative skills. Omnibus differences were reported for Positive Behavioral Expectations at school level in school 3, Lesson Facilitation in school 3 and also for Perseverance in all the schools and Social and Collaborative Skills in school 1, 3 and 4. There were no omnibus differences within schools for Supportive Learning Environment, Checks for Understanding, Feedback, Critical Thinking, and Autonomy. The results are consistent with improvement with repeated feedback cycles, but the lack of control group, short observation period and extensive pairwise testing make it difficult to draw causal conclusions. One aspect of professional development that might be fruitful is focused feedback from the observer, which can be repeated and translated into the school context.

Keywords: Observer Feedback; Teaching Effectiveness; Classroom Observation; Professional Development; Science Education; Mathematics Education; Grade VI

1. Introduction

Teaching The quality of teaching is a crucial element of school effectiveness that impacts on pupil learning, practice and recall. Explanation, questioning, monitoring of student understanding, and classroom interaction are particularly significant in science and mathematics to help students

advance beyond recall to reasoning and problem solving, and to develop conceptual understanding. Educationally the Grade VIII stage is very important as learners are expected to be more independent, and deal with more complex content, yet still need clear structure and teacher support. Therefore, there is a very practical dimension to professional development that is very close to classroom practice and is able to be applied in the school to enhance the quality of teaching and learning. A direct connection between professional development and practice can be done through observer feedback. An observer may document specific classroom behaviours, strengths and areas for change and discuss with the teacher rather than just workshops or general advice. Feedback research highlights the importance of specificity, link to desired standard, actionable information (Hattie & Timperley, 2007). Professional development research also confirms that ongoing learning and linkages to classroom content and issues are more likely to lead to changes in practice (Garet et al., 2001). It is also supported by feedback from diverse sources which has been associated with changes in practice in teacher mentoring when the mentoring is structured and supportive (Keiler et al., 2020). Classroom observation can enhance this process as this shifts the focus to what is really happening in the classroom (Anwar & Menekse, 2021). Assessment of the quality of interaction has been correlated with engagement and achievement in mathematics classrooms, thus providing a model of the importance of systematic focus on classroom processes (Wang et al., 2020). For while, seeing is not seeing objectively. Ratings can be influenced by the observation window, the relationship between the observer and the teacher, and the training and judgement of the observer (Marsh, 2007; Munson, 1998). It is important that a good feedback system have standardised criteria, repeat observations and careful interpretation.

The case study study for the thesis explored how the feedback of the observer was being used by the grade VIII science and math teachers in four public school of Lahore. There was repeated observation of teachers with feedback following each cycle of observation. The original thesis comprised of four research questions, two of which related to teacher perceptions, and the quality of the feedback. A separate perception scale, interview schedule or feedback quality instrument was not used to assess these constructs. The following questions have been selected to focus the present manuscript on questions that can be answered, based on the observation scores recorded over the five feedback cycles: Do observed teaching scores improve over five feedback cycles? Are there similar trends in growth in observed teaching scores across schools? What aspects of teaching/learning demonstrate the greatest improvement? This is a more focused analytical lens that will increase the construct validity and not attribute perceptions that the researcher did not measure on the respondents.

The article therefore shares observations of the changes occurring in nine aspects of teaching, grouped in Classroom Culture, Instruction and Socioemotional Skills. It isn't a cause, it's a contribution. It provides a short story of changes to specific teaching practices based on a series of short systematic observations and feedback, and shows that not all domains or schools necessarily followed the pattern of change. This is important for school leaders because while a feedback process may have a positive effect even if improvement is not even, the findings should be a diagnosis and based on targeted support and not a performance score.

2. Literature Review and Conceptual Basis

Feedback is generally considered to be information that assists an individual in comparing his/her level of performance to the desired level of performance and what is expected to change next. Hattie and Timperley (2007) considered effective feedback to be in the direction of learning, the

quality of the learning that was occurring and what was needed to do better. The same logic is used to interpret teacher learning when feedback is related to observable classroom practice, although their framework is often used for student learning. Akkuzu (2014) also concluded that there are different type of feedback which can interact with the teacher's performance and teacher self-efficacy, which means that feedback can influence the teacher's performance and teacher self-efficacy as teacher.

Feedback is magnified where professional learning is continuous and integrated into practice. Garet et al. (2001) found that subject matter focus, time, and active learning opportunities were more likely to result in professional learning that impacted teaching. Keiler et al. (2020) showed that mentoring and feedback structured and connected with the model being built in the classroom, can be of support in the change process of the teacher. Van Es and Sherin (2008) also demonstrated that with the help of a guide to attend to classroom events, mathematics teachers can see some aspects that may not have been seen otherwise. These studies in combination provide for a cyclical observation process that includes feedback and an opportunity for reenacting and reassessing the change.

Good teaching has many aspects. Stronge et al. (2011) related teacher effectiveness to instructional delivery, assessment practices, learning environment, and personal and relational qualities. The three broad domains given in the present study are in accordance with these dimensions. Classroom Culture is a representation of the positive climate and expectations of teacher behaviour. Instruction includes lesson facilitation, checks for understanding, feedback to the student, and critical thinking. Socioemotional Skills are student autonomy, perseverance, social and collaborative skills. So, teaching effectiveness is seen as a pattern of observable practices and not a score.

In literature, improvement is also stated to have to be understood in light of context. It is a fact that teacher effectiveness differs from teacher to teacher in the same school (Nye et al., 2004) and that through collaboration, learning, and networking with each other, teachers can impact professional practice over time (Jackson & Bruegmann, 2009). Thus, differences between schools may be more than the difference in the feedback message. The level of implementation of feedback can vary depending on teacher experience, resources, students' composition, norms in the profession, and leadership support. This is particularly true if the interventions are carried out in a limited number of schools and there is no control group.

The following two notions of concepts prove useful in the current study. Feedback theory can help us to understand the mechanism by which repeated, and specific, information concerning classroom practice may help teachers realize that there is a gap between what they are doing and what they wish to be doing. Second, from a social learning and self-efficacy perspective, teachers will be more likely to adopt new practices if they can see the standards in action, if they are given credible information about their performance, and if they are able to make reasonable improvements over time (Bandura, 1997). The findings are interpreted with the help of these perspectives. They did not decide on the design of the observation instrument used; it was a preexisting structured instrument.

3. Method

3.1 Design and Setting

The study was a single group quasi experimental study with repeated classroom observation and feedback. It was conducted in 4 public schools of Lahore and addressed to grade VIII science and math teachers. No comparison or control group data. The design is therefore appropriate for describing changes that happen between one observing position and another and that take place over time, but it is not suitable to link changes only to the feedback of the observer.

3.2 Participants and Sampling

The population of the study is science and mathematics teachers having public schools of Punjab and sample is limited to Grade VIII teachers of four selected schools of Lahore. Convenience sampling was used as a sampling method. The outputs for the school specific means and paired t tests are for 28 teachers in the analytic sample: eight teachers in each of Schools 1, 2 and 3, and four teachers in School 4. The teacher demographic characteristics were also gathered on the instrument but were not reported in the results of the thesis and are not reconstructed or introduced into this article.

3.3 Observation Instrument

The Teach Observer Manual 2nd Edition Observation Sheet was used to collect data. It contained a demographic section and a closed-ended classroom observation section that consisted of 28 statements in 9 elements in 3 domains. Classroom Culture was broken down into Supportive Learning Environment and Positive Behavioral Expectations. Lesson Facilitation, Checks for Understanding, Feedback, and Critical Thinking were taught. The Socioemotional Skills categories were Autonomy, Perseverance and Social and Collaborative Skills. In the statement scores are converted to 5-point rating system, according to the thesis. The nine elements each have varying numbers of underlying statements so that raw totals of elements are interpreted over time within each element, and not as individual effect sizes across elements. A pilot study was conducted in a school not sampled and items on the instruments were discussed with experts before data collection was done. The thesis states that the internal consistency was tested through Cronbach's alpha, however the coefficient is not presented. Hence, for the present article, the pilot and expert review procedures reported are kept without claiming any reliability value.

3.4 Procedure

Classroom observation was carried out after gaining permission from the principals and cooperating teachers. The same teachers were observed on five observation points with a gap of 1–2 days between each. The thesis states that the Teach procedure consisted of two 15-minute lesson observation segments. After observing, the checklist was completed to find out what needed to be improved and the researcher consulted with the teacher regarding the items on the checklist that were observed, typically during a break or a free period. This observation, scoring and feedback was then repeated in the cycle.

3.5 Statistical Analysis

The thesis analysis is carried out in SPSS version 25. For each of the five observation points for each teaching element, the means and standard deviations were calculated. Pairwise comparisons between observations were chosen and carried out using paired sample t-tests, and analysis of variance was provided for school level comparisons over the sequence of observations. The .05

significance level was used to determine statistical significance. In this article, the results of the inferential tests included in the thesis are interpreted in an exploratory fashion and not as definitive estimates of intervention effect because of the lack of a multiplicity correction and the number of pairwise tests conducted.

3.6 Ethical Considerations

The thesis states that permission to carry out the work was given by the research supervisor and principals and relevant teachers were contacted before the data was collected. They were told that they could choose to participate or not and that the data would be used for research and be confidential.

4. Results

For the whole sample, the mean score for each of the nine Teaching elements observed at Observation 5 showed an improvement over scores from Observation 1. However, sometimes the trajectory's paths were not straight lines. There were at least one interim decrease in Supportive Learning Environment, Lesson Facilitation, Feedback, Critical Thinking and Autonomy and more steady upward increases in Positive Behavioral Expectations, Checks for Understanding, Perseverance and Social and Collaborative Skills. Domain below provides the descriptive results with details.

4.1 Classroom Culture

Element	Obs. 1	Obs. 2	Obs. 3	Obs. 4	Obs. 5
Supportive Learning Environment	11.53 (3.36)	10.46 (3.69)	10.89 (3.55)	11.85 (2.97)	12.17 (2.89)
Positive Behavioral Expectations	9.53 (2.20)	10.17 (3.05)	10.50 (3.41)	10.96 (2.48)	11.46 (2.38)

Note. Values are mean scores with standard deviations in parentheses. Data reproduced from the thesis Table 4.1.

Supportive Learning Environment started off at 11.53 at first observation, and decreased to 10.46 at second observation before slowly improving to 12.17 at fifth observation. Positive Behavioral Expectations improved in a more predictable manner, from 9.53 to 11.46. The descriptive increase was modest (1st – 5th) for Supportive Learning Environment and more prominent (1st – 5th) for Positive Behavioral Expectations. The omnibus ANOVA at the school level revealed no significant difference between the observation points for Supportive Learning Environment in each of the four schools. There was only a significant omnibus difference for Positive Behavioral Expectations in School 3, $F = 5.97$, $p < .05$.

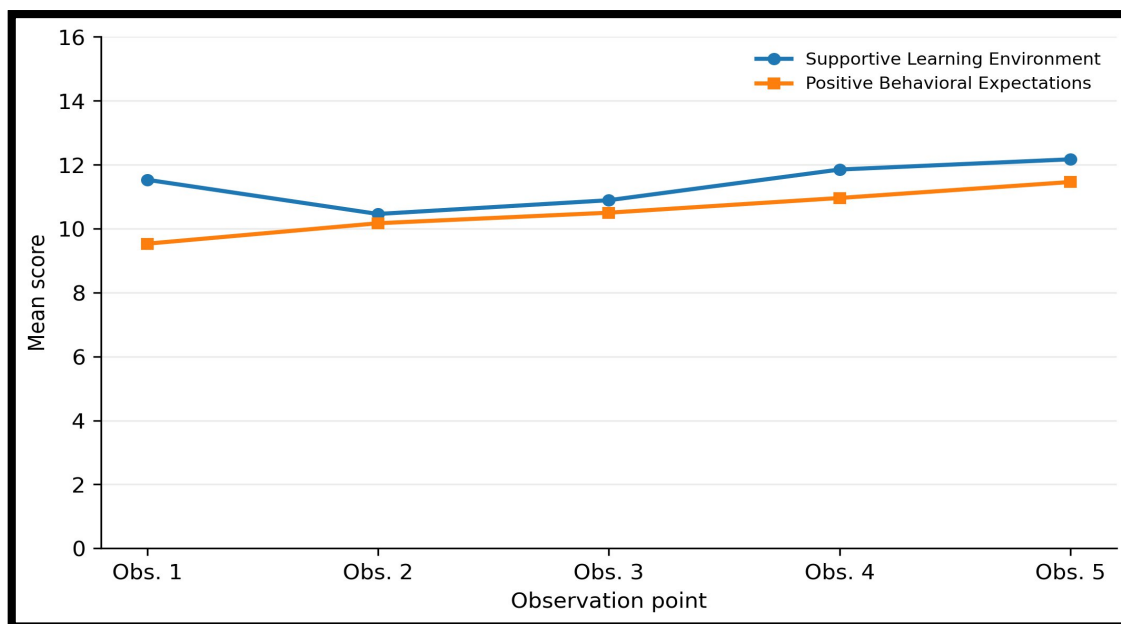


Figure 1. Classroom Culture mean scores across five observations. Values reproduce the thesis results.

4.2 Instruction

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 Understanding	10.28 (2.78)	10.42 (3.29)	11.14 (2.95)	11.85 (2.25)	12.00 (1.58)
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 mean scores, standard deviations are shown in parentheses. These data have been replicated from the thesis Table 4.2.

The scores of the Lesson Facilitation clearly increased over time (Observation 1: 12.17; Observation 2: 11.74; Observation 3: 13.16; Observation 4: 14.32; Observation 5: 15.10). Checks for Understanding slowly increased from 10.28 to 12.00. Feedback was the least stable element of instruction with a mean score of 6.21 at Observation 2 that increased to 7.96 at Observation 4 and slightly decreased to 7.64 at Observation 5. Overall, Critical Thinking improved from 10.57 to 12.28, although there were minor drops between Observations 2 and 3, and between Observations 4 and 5.

Lesson Facilitation was significantly different on an omnibus basis across the five observation points in School 3, $F = 5.77$, $p < .05$. There were no omnibus differences in any school for Checks for Understanding, Feedback or Critical Thinking. The pairwise analyses for these elements revealed some significant differences, but due to the number of pairwise tests that were conducted, without multiple comparisons adjustments, the following pairwise differences are reported with caution.

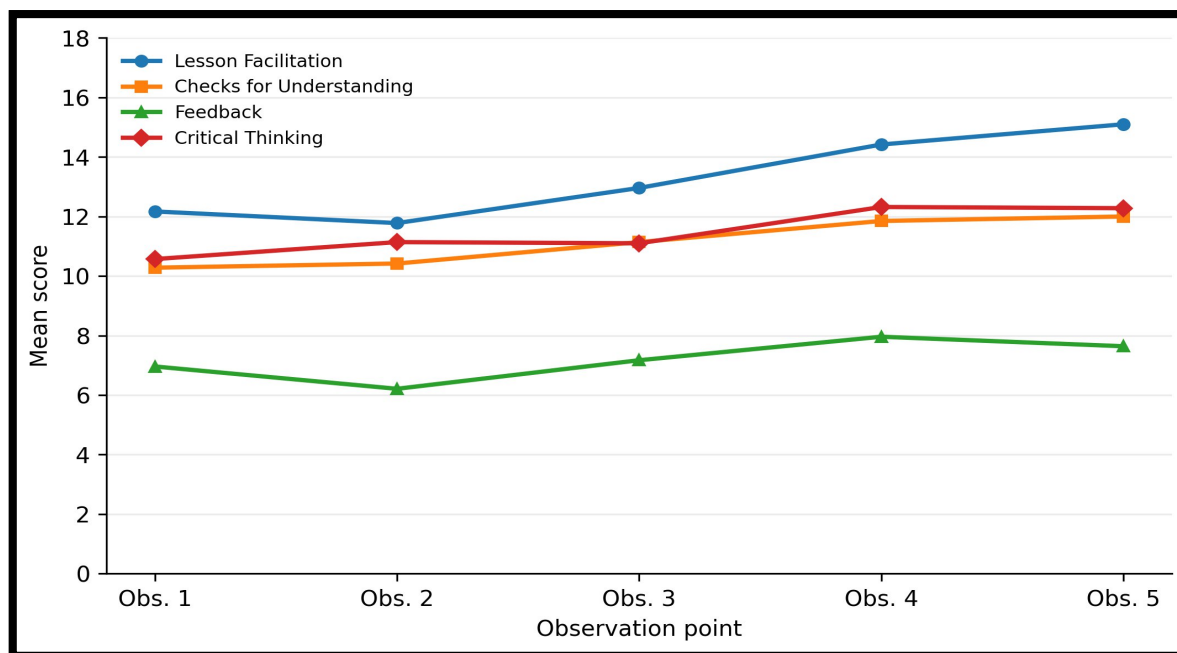


Figure 2. Instruction mean scores across five observations. Values reproduce the thesis results.

4.3 Socioemotional Skills

Element	Obs. 1	Obs. 2	Obs. 3	Obs. 4	Obs. 5
Autonomy	8.96 (3.19)	10.39 (3.42)	9.57 (3.50)	12.14 (2.39)	11.07 (2.77)
Perseverance	7.28 (2.80)	7.67 (2.89)	7.92 (3.79)	10.57 (2.82)	11.50 (3.39)
Social and Collaborative Skills	7.25 (4.35)	8.03 (3.92)	9.10 (4.02)	11.17 (3.68)	12.10 (3.76)

Note. The values are presented as mean scores (standard deviations). Data on the table is from the thesis Table 4.3.

The autonomy rose from 8.96 at first observation to 10.39 at second observation, dropped to 9.57 at third observation, reached a peak of 12.14 at fourth observation and reached a value of 11.07 at the end of the observations. For its fifth observation, Perseverance was at 11.50 with a much more positive upward trajectory, from 7.28. Social and Collaborative Skills also steadily increased from 7.25 to 12.10. The socioemotional trajectories indicated that there were considerable changes in these domains during the time of observation, although the trajectory for Autonomy was not as consistent as the trajectory for Perseverance and Social and Collaborative Skills.

These social and emotional aspects of student performance had the best scores for the school level. Perseverance differed significantly across the five observations in all four schools: School 1, $F = 14.23$, $p < .012$; School 2, $F = 8.13$, $p < .03$; School 3, $F = 28.86$, $p < .003$; and School 4, $F = 8.13$, $p < .03$. Social and Collaborative Skills differed significantly in School 1, $F = 11.93$, $p < .01$; School 3, $F = 15.21$, $p < .01$; and School 4, $F = 11.93$, $p < .01$, but not in School 2, $F = .53$, $p = .72$. There were descriptive increases over time, but no significant differences in increase in autonomy by school.

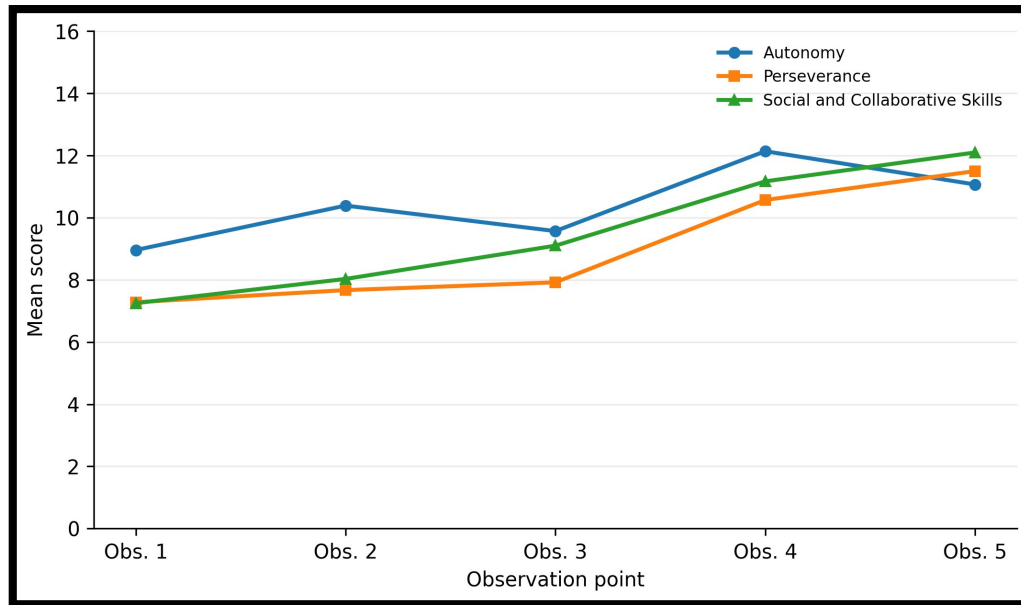


Figure 3. Socioemotional Skills mean scores across five observations. Values reproduce the thesis results.

4.4 Summary of School Level Omnibus Results

Teaching element	School	F	p
Positive Behavioral Expectations	School 3	5.97	< .05
Lesson Facilitation	School 3	5.77	< .05
Perseverance	School 1	14.23	< .012
Perseverance	School 2	8.13	< .03
Perseverance	School 3	28.86	< .003
Perseverance	School 4	8.13	< .03
Social and Collaborative Skills	School 1	11.93	< .01
Social and Collaborative Skills	School 3	15.21	< .01
Social and Collaborative Skills	School 4	11.93	< .01

These social and emotional aspects of student performance had the best scores for the school level. Perseverance differed significantly across the five observations in all four schools: School 1, $F = 14.23$, $p < .012$; School 2, $F = 8.13$, $p < .03$; School 3, $F = 28.86$, $p < .003$; and School 4, $F = 8.13$, $p < .03$. Social and Collaborative Skills differed significantly in School 1, $F = 11.93$, $p < .01$; School 3, $F = 15.21$, $p < .01$; and School 4, $F = 11.93$, $p < .01$, but not in School 2, $F = .53$, $p = .72$. There were descriptive increases over time, but no significant differences in increase in autonomy by school.

There were no significant omnibus differences for Supportive Learning Environment, Checks for Understanding, Feedback, Critical Thinking, or Autonomy. The omnibus pattern is the elaboration of the descriptive picture. There were within time differences that were statistically significant in Positive Behavioral Expectations, Lesson Facilitation, Perseverance and Social and Collaborative Skills. Perseverance was the element that was significant in all four schools, and repeatedly in the

significant results in school 3. This distinction is significant because there need not be any necessary or required relationship between descriptive improvement and statistical significance.

5. Discussion

The descriptive signal (the study) is consistent, showing that observed teaching practice was enhanced with repeated cycles of observation and feedback. The mean score for each measure rose at the 5th observation compared to the 1st observation. The pattern was most apparent descriptively for dimensions of Perseverance, Social and Collaborative Skills, and Lesson Facilitation, and Feedback and Supportive Learning Environment showed less change. The findings support the notion that multiple feedback facilitates teachers to focus on particular practices; however, there is no control group, thus the findings cannot be interpreted as proof of the impact of feedback alone in facilitating the changes.

This improvement applies in particular for the feedback literature to Lesson Facilitation. Hattie and Timperley (2007) claim that the feedback that is most helpful is the feedback that 'makes sense' of the current performance and what to do next. The lesson facilitation is transparent and observable: that is, an observer can observe how the teacher is organizing the lesson, how the teacher is explaining content, how he or she is supporting participation, how he or she is managing the transitions, etc., while at the same time the teacher can make a concrete change in the next lesson. This could account for the clear descriptive increase and significant school level omnibus result for Lesson Facilitation in School 3. The finding is also consistent with the results of professional development research that indicates that developing learning that is directly connected to classroom content and practice is more likely to impact classroom practice than is generic development activity (Garet et al., 2001).

In addition, there was also a steady trend toward higher Positive Behavioral Expectations at the full sample level and a significant omnibus result in School 3. This is visible Classroom Culture - one part that can be communicated to teachers. According to Stronge et al (2011) an effective teacher provides an environment and expectations, in addition to delivering the content. Therefore, it might be helpful to repeat the observation to see if expectations are set and revisited, or simply discussed as a management principle. Evidence of school level is most evident for Perseverance and Social and Collaborative Skills. Perseverance and Social/Collaborative Skills had wide range of variance across the observation sequence in all four schools and in three of the four schools, respectively. These findings do not only concern the mechanics of the instruction, but also the relevance of observers' feedback. They suggest that this observation process can be applied to draw teachers' attention to classroom activities that help pupils to work hard and interact and participate with peers. Similarly, Wang et al. (2020) show that classroom interaction quality is important for mathematics, as is the importance of learning through interaction among professionals, as demonstrated by Jackson and Bruegmann (2009). There is no direct relationship to the present results and student achievement, but the results suggest that some aspects of SEL in the classroom were responsive during the time period observed.

In contrast, the Subscale of Supportive Learning Environment, Checks for Understanding, Feedback, Critical Thinking and Autonomy did not show significant school level differences in their overall

means while the overall means increased. This difference between descriptive and inferential evidence is important. There will be some teaching practices that will not be stabilised in five cycles of feedback; some will be more sensitive to class composition, content of the lesson or some will be more difficult to capture in five cycles of observation. For example, feedback to students and checks for understanding may be rapid, and may occur in multiple forms throughout a class, and then they would be responsive to the timing of the observer's visit. Critical thinking may also be related to the specific task being taught on a specific day. Even though these explanations are viable, they are not findings as the thesis did not gather contextual qualitative data to test these explanations.

School context also was a factor. School 3 was significantly different than the other schools on omnibus Positive Behavioral Expectations, Lesson Facilitation, Perseverance, and Social and Collaborative Skills; School 2 was significantly different only on Perseverance. This should be interpreted as a general trend and not a "top to bottom" ranking of schools. The thesis did not include any analysis of teacher experience, leadership support, timetabling, classroom conditions or student characteristics and therefore the reasons for differences between schools are not clear. The finding, however, represents a context understanding of feedback. A regularised feedback process can be set up across schools but its uptake and the results seen will depend on the local situation. The article also clarifies one aspect of measurement not so explicitly addressed in the thesis. The original research question(s) focused on teacher perceptions of the feedback and the quality of feedback, but the instrument reported did not focus on teacher perceptions or feedback quality. An observation score will not be used to answer a perception question unless the question is of a high quality. It is therefore only possible to make conclusions for the observed teaching practices in the present manuscript. Future research can therefore include the original perception questions, however, in addition to classroom observation, interviews or questionnaires, or a validated feedback quality measure is needed.

5.1 Practical Implications

The findings provide school leaders with an understanding of how to work with the observation process as a developmental process and not an isolated inspection. The study design provided several times to offer feedback to the teachers and to make another observation. This is consistent with research that indicates that ongoing professional learning, specific to a job, and in practice, is more effective (Garet et al., 2001; Keiler et al., 2020). Then, 1 or 2 priorities would be identified at the end of each visit, some feedback provided and a follow-up visit conducted to those priorities.

The findings also suggested that there should be a greater emphasis on some socioemotional aspects of classroom practice in feedback procedures in addition to content coverage. The school level omnibus had the best evidence for Perseverance and Social and Collaborative Skills. For example, the presence of student persistence, opportunities for collaboration and level of productive autonomy provided during the lesson could be included in the supervisor's promptings. In the meantime, those that did not change as frequently should not be forgotten. Feedback, Critical Thinking and Checks for Understanding may require more time and/or a variety of sources of evidence and observation time to be larger or longer.

Finally, observation systems should make every effort to separate developmental feedback from high stakes judgement, when possible. Rating on observation is influenced by context and interpretation of the observer (Marsh, 2007; Munson, 1998). The best use of the current results is, therefore, as a tool for coaching, reflection, and targeted professional development, rather than as a definitive indicator of employment or accountability based on a brief observation score.

5.2 Strengths and Limitations

The repeated observation structure is a strength of the study. The measurements of the same teaching elements at 5 time points also allowed the trajectories to be illustrated in the article and not just a pre and post comparison. The use of a structured observation instrument also gave a common framework in the four schools and enabled change in specific elements of teaching to be reported, instead of a single, overall judgement. There are some restrictions which have a material bearing on the interpretation. The changes found in this study may have been due to maturation, familiarity with the instrument, the effect of the observer, or other school activities, as there was no control group. Second, the sample was obtained from four public schools of Lahore, thus limiting the generalisability of the sample. Third, two observation periods, each 15 minutes long, are reported in the thesis, and these might not cover the spectrum of practice for a whole lesson or teaching day. Fourth, teacher demographic data were not provided, and therefore could not be analysed by experience, subject specialism, gender and other factors. Finally, Cronbach alpha was not reported despite the fact that, as mentioned in the thesis, reliability was measured. Sixth, effect sizes and confidence intervals were not reported. Seventh, a number of pairwise t tests were not corrected for multiple testing and thus the possibility for spuriously significant results exists. In this regard, the present article is more interested in patterns of description and omnibus school level results.

Also, no direct measure of student learning results or teacher perceptions was used. So, it can't be concluded that there was a link between better observation marks and better attainment in mathematics or science, nor can it say anything about the teachers' experience of the feedback process. These are important fields in which more research can be conducted.

6. Conclusion

In this article, the thesis data were revisited to give a more condensed and publishable report on the feedback and effectiveness of teaching of the Grade VIII science and mathematics teachers. Mean scores on all nine teaching elements at all five observation points were higher at the end of the measurement than at baseline, but not necessarily. There were less consistent patterns of growth or no noticeable patterns of growth in Checks for Understanding, Feedback, Critical Thinking and Autonomy, while Supportive Learning Environment, Perseverance, and Social and Collaborative Skills showed the most consistent evidence of growth.

School level analysis shows that context is important. The omnibus changes in School 3 were significant on several elements and the omnibus changes for perseverance were significant at all four schools, and for Social and Collaborative Skills were significant at three schools. These outputs also provide evidence for the effectiveness of repeated structured observer feedback as a professional development process, including the importance of the specificity of feedback, another opportunity to

practise and feedback understood within the context of local conditions in schools. But it doesn't warrant causal claims about the effects of feedback for the design. A comparison group, longer follow up, effect size, and confidence intervals, and a predetermined model for multiple testing, and direct assessment of student learning and qualitative teacher perceptions and feedback received could be considered for future research.

The main implication is therefore practical but important: the repeated observations and feedback allows for a structured way of identifying change in classroom practice and thus, providing professional support. It has its most significant contribution in development. It can be a valuable tool to move from general supervision to evidence for schools to increase the quality of science and mathematics education, when integrated with trained observers and clear criteria and follow up coaching.

References

- Akkuzu, N. (2014). The role of different types of feedback in the reciprocal interaction of teaching performance and self efficacy belief. *Australian Journal of Teacher Education*, 39(3), 37–66.
- Anwar, S., & Menekse, M. (2021). A systematic review of observation protocols used in postsecondary STEM classrooms. *Review of Education*, 9(1), 81–120.
- Bandura, A. (1997). *Self efficacy: The exercise of control*. W. H. Freeman.
- 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.
- Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, 1(4), 85–108.
- 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.
- Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In *The scholarship of teaching and learning in higher education: An evidence based perspective* (pp. 319–383). Springer.
- Munson, B. R. (1998). Peers observing peers: The better way to observe teachers. *Contemporary Education*, 69(2), 108.
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26(3), 237–257.
- Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education*, 62(4), 339–355.
- 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.
- Wang, M. T., Hofkens, T., & Ye, F. (2020). Classroom quality and adolescent student engagement and performance in mathematics: A multi method and multi informant approach. *Journal of Youth and Adolescence*, 49, 1987–2002.