

TRAINING NEED ANALYSIS AND ITS EFFICACY FOR TEACHERS' EVALUATION

Muhammad Faheem Asher¹, Syed Mohsin Ali Shah²

Abstract

In many schools, teacher evaluation is only used to provide teachers a score or a category. It is not employed as a tool in assisting teachers to grow. This paper looks at the relationship between teacher evaluation and Training Needs Analysis (TNA). The main argument of this research is that a better evaluation system, one that provides proper feedback, is needed before any real TNA can happen. This paper uses a case study of a private chain of schools and colleges in the North Zone of Pakistan. In this case study, twenty teachers were observed in real lessons. Thirty teachers answered a Likert-scale survey, and some heads of section were interviewed. Along with this case study, the paper also utilizes international research articles on teacher evaluation, needs analysis, and classroom observation. The paper shows how evaluation data can highlight the strong and weak areas of teachers. Further, it also explores how good feedback can turn these findings into professional growth, and how schools use evaluation results to improve their training programs. The results show a clear pattern of strengths and weaknesses. For instance, teachers were strong in subject knowledge, using technology and communication. On the other hand, they were weak in lesson planning, writing learning objectives, using different teaching methods, and asking good questions. The similar pattern is also seen in other education systems. The paper connects these results to the Three-Level Training Needs Analysis (TNA) Model, and argues that schools should make proper de-briefing, evaluation-based TNA cycles, and mentoring a normal part of teacher training.

Article History

Received 15 April 2026

Revised 25 June 2026

Accepted 24 June 2026

Published 30 June 2026

Corresponding Author*
OPEN ACCESS

Keywords

Teacher Evaluation,
Training Needs Analysis,
Constructive Feedback,
Pedagogical Skills,
Professional Development.

1. Introduction

1.1 Background and Rationale

The most common exploration in education research is that teacher quality is the biggest school-related factor that affects student learning (Darling-Hammond, 2000; Hattie, 2009). Factors like curriculum, class size and school building matter as well, but nothing has as much effect as the classroom teacher. Resultantly, the systems that schools use to evaluate and improve teacher quality through feedback mechanism become very important. Teacher evaluation is one of the main tools used to achieve purpose.

However, teacher evaluation, both in Pakistan and around the world, has always faced challenges. It tries to do two things simultaneously that do not go hand in hand: development and accountability. Development means finding out what a teacher needs to improve and using that as a base for training. Accountability means evaluating if a teacher is good enough to keep the job, get promotion, and in which category to be placed. (Toch & Rothman, 2008) and (Weisberg et al.,

¹ ssistant Professor, College of Education, PAF. faheemasher123@gmail.com

² B.A (Hons). in English Literature, B.Ed..Education Instructor, College of Education, PAF
syedmohsin2777@gmail.com

2009), both found that evaluation systems in America in the early 2000s were not good at either of these jobs. They could not differentiate between good and weak teachers, and even when they could, this information was merely not used to help teachers grow. This is not only a particular country related problem. However, it happens anywhere evaluation is treated as a final step instead of the first step of a process.

Training Needs Analysis (TNA) originates from the field of human resource management, and it can help resolve this problem. (McGehee & Thayer, 1961), were the first to explain TNA using three levels: organizational, person, and task. TNA is basically a way to differentiate between how people perform now and what they should be able to do after receiving feedback. When we put this approach in practicality, TNA asks a simple question: what do these teachers need to learn, and how do we know this? Teacher evaluation, if it is based on clear objectives and followed by proper feedback, it could be one of the few tools that can address this question using real evidence, not just guesses.

1.2 The Pakistani Context

In Pakistan, this issue is even more of importance. Data from the National Education Census show that a lot of teachers in the private school sector do not possess a formal teaching certificate. The National Professional Standards for Teachers in Pakistan (2009) were made to promote an idea of what a good teacher should be able to do, across ten areas, most importantly subject knowledge and use of ICT. However, after the 18th Amendment gave education powers to the provinces in 2010, the focus on teacher training across provinces became weaker. There is no single national system now that keep check and balance whether in-service training programs meet a proper standard or not. Some studies also show that these standards are not properly followed. A comprehensive study on secondary teachers in Lahore found that only about half of the teachers used differentiated instruction as required by the standards, and the weak areas were assessment and lesson planning (Investigation of Teachers' Practices through National Professional Standards for Teachers in Pakistan, n.d.). Because the system is not very organized and policies not implemented. Many school chains design their own way of connecting evaluation with training, which makes case studies like this one useful.

1.3 Statement of the Problem

In most Pakistani schools, teacher evaluation is only used to categorize teachers, and then nothing more happens. The data collected during evaluation program is merely used properly to find out what sort of training the whole institution needs. Teachers are not provided with a clear path to turn their evaluation results into a personal development plan. This paper intends to fill this gap. It asks questions like: whether, and how, teacher evaluation can work as a first step before Training Needs Analysis in a real school setting. With the help of a case study of a private chain of schools and colleges as the main example, and connects it to broader research from other countries.

1.4 Objectives

This paper has three main objectives:

1. To explore how teacher evaluation systems, work generally in schools and particularly in a Pakistani private-school chain. Further, to highlight the strengths and weaknesses of teachers.
2. To evaluate the role of good, constructive feedback provided during de-briefing to help teachers grow professionally.
3. To look at how decisions about training program design, in-service teacher training, and mentoring are affected by the evaluation outcomes.

1.5 Research Questions

1. What are the areas commonly used to evaluate teachers, and in the case study which of these areas are strong or weak?
2. What is the role of constructive feedback in helping teachers use their evaluation results?
3. In what ways does the evaluation-based Training Needs Analysis affect design of in-service training programs?

2. Literature Review

2.1 What Teacher Evaluation Is, and What It Is For

In simple words, evaluation is a process of judging something. Data is collected based on some standard, and this data addresses one question: how good this practice is, and if it meets the expected standard? Teacher evaluation applies this similar idea to classroom teaching. Over time, the standards used to evaluate teachers have changed a lot. According to Jewell this started with inspection models in the 19th century, where a teacher was considered as good if the required curriculum was taught, classroom looked organized, and the discipline was kept. Later, in the 20th century, supervision models came, and now the focus shifted towards the value-added and standards-based evaluation. It connects teaching directly to student results (Jewell, n.d.). Currently, the idea of teacher quality is much broader than the old checklist. It includes subject knowledge, relationship with students, use of assessment, classroom environment, high expectations, ability to teach diverse learners, continuing professional learning, and teamwork with colleagues and parents.

Teacher evaluation should not be seen simply as a way to determine a performance rating, but can be a systematic process to identify strengths, weaknesses and areas for professional growth. Robinson (2022) argues that the aim of teacher evaluation is not merely accountability but to improve teaching practice and consequently, student learning. However, the value of evaluation is only as good as the evidence gathered, the criteria employed, and the extent to which evaluation findings are translated into meaningful professional learning opportunities. This is especially important for Training Needs Analysis (TNA) as a summative score only evaluation provides very little information as to what a teacher needs to learn or improve. Conversely, evaluation systems that produce specific and actionable evidence can be used to identify competency gaps and develop specific training interventions.

The movement toward teacher evaluation systems also reflects a broader shift from relatively simple inspection and compliance-oriented approaches toward more systematic attempts to assess teaching quality and improve instructional practice. Rucinski and Diersing (2014) describe the evolution of teacher evaluation in the United States within the framework of a broader reform movement in which evaluation increasingly became associated with accountability, standardized

measures, and attempts to differentiate teacher performance. This evolution matters for understanding the contemporary debate about teacher evaluation: although more systematic evaluation can generate evidence of teaching quality, it becomes educationally meaningful only when that evidence is used to support teacher learning and professional improvement. The value of an evaluation system, therefore, should be measured not only by its ability to classify teachers, but also by its ability to identify areas for development and to link those areas to appropriate professional learning.

Every evaluation system often tries to do two things simultaneously: development and accountability. Development treats evaluation as a coaching tool to improve teaching over time. Some well-known frameworks show this difference clearly. Accountability treats evaluation as a tool for personnel decisions, such as keeping a job, giving tenure, or deciding pay. Formative evaluation is low-stakes and more focused on improvement. On the other hand, summative evaluation, like Danielson's Framework for Teaching, is higher-stakes and connected to standards. It can affect contract decisions, even though its results can also be used for the purpose of professional growth (Danielson, 2013). Research on classroom observation highlights that these two purposes need different levels of reliability. A small number of observation visits is enough for useful formative feedback, but high-stakes decisions. For instance, decisions about tenure, which need much more observation done by more than one observer (Center for Faculty Development and Excellence, n.d.). A single lesson observation cannot give concluding remarks about a teacher's overall performance. However, it can provide useful feedback about that one lesson. This point is important for this paper's argument. An evaluation system that is primarily built for accountability is not good at providing the kind of detailed and area-specific information that TNA needs. A system that is mainly built for formative feedback works much better for this purpose.

2.2 Teacher Quality, Feedback, and Student Learning

The reason teacher evaluation is worth it is because of the strong connection between teacher quality and student achievement. A well-known study of state policies in the United States by (Darling-Hammond, 2000) found that teacher preparation and certification were more strongly connected to student achievement in reading and math than any other school factor, including class size and spending per student. Later research also emphasizes teacher-related factors, including feedback quality, among the strongest school-based influences on student learning (Hattie, 2009). If teacher quality matters this much, then the systems used to find and fix gaps in teaching become very important as well. Because these systems are the primary way a school can improve its existing teachers, not just through inducting new ones.

Feedback is the integral part that decides whether evaluation actually helps teachers grow. (Tuytens & Devos, 2017), in their research study of eight schools, found that whether teacher evaluation led to real improvement depended less on the evaluation tool itself, and more on how the school was organized. Most importantly, whether school leaders created a two-way feedback process during the whole evaluation cycle or treated feedback as a one-time formality at the end. When school leadership did not do this job, evaluation results were mostly left unused. A second study of feedback from classroom observation suggests that there are two conditions that need to be met for feedback to be useful. Feedback needs to be based on a clear, shared standard so that teachers and observers are 'on the same page'. This way the teacher knows exactly what to change, fast and

hopefully repeated over several cycles rather than a one off event (ClassMind: Scaling Classroom Observation, 2025). Feedback that ignores these conditions, which is vague, is not provided on time, or is given only once, is often considered by teachers as just another compliance task, and this reduces its usefulness for TNA.

The connection between evaluation and professional development is most apparent when feedback points to specific instructional needs. As an example of this relationship, Nausheen et al. (2024) used a faculty-development model that incorporated traditional needs assessment with student evaluations and performance appraisal. Feedback was found to be able to identify specific needs for development in areas like teaching skills, assessment practices, curriculum design and educational technology. Crucially, the evaluation was not considered a final destination, but the feedback was used to change and expand faculty development activities, establishing a continuous cycle of assessment, training, feedback, and further development. This is empirical evidence for the view of teacher evaluation as a possible data source for TNA, not only performance judgment. In this way, evaluation evidence can help organizations move from the general knowledge that teachers need professional development to the more specific determination of what teachers need to learn, why they need to learn it, and which training interventions are most appropriate.

It is argued that not all research agrees that feedback always leads to growth. (Derrington & Martinez, 2019) in their research studied a new evaluation system used in nine school districts in eastern Tennessee for five years. They came to know that the new system changed the relationship between head of the institute and teachers and took more time, but it did not clearly provide more growth chances to teachers who were already rated as competent. This shows that we should not assume that every new evaluation system automatically leads to development. (Jarocki, 2018) did the survey high-school teacher trainers and administrators in California and found that student evaluations of teaching were considered as useful for professional development, however, some teachers still resisted this idea. This reveals to us that whether evaluation data actually helps depends a lot on how it is introduced and who controls it.

2.3 Training Needs Analysis: Concept and Models

The term Training Needs Analysis actually comes from organizational psychology and human resource development, not from education (McGehee & Thayer, 1961). First explained that training needs should be checked at three levels. First, the organizational level, which asks if training is really the right solution for a bigger problem in the whole organization. Second, the task level, which explains what knowledge and skills a certain job actually needs. Third, the person level, which measures the gap between what a person can do now and what the task level says they should be able to do. Later researchers, like Ostroff and Ford, contributed more detail to this model but kept the same three-level idea (Models – NeedsAssessment.org, n.d.). This O-T-P model is still popular because it stops training decisions from being made without looking at the bigger picture. A need found only at the person level, without checking what the job needs or why the organization should invest in it, can lead to training that looks better but is not really useful.

A further strength of this approach is that training needs need not be seen as fixed or identified thru a single assessment exercise. Nausheen et al. (2024) illustrate a cyclical faculty-development

process where initial needs assessment informed training, while subsequent student evaluations generated further evidence used to refine the development program. This approach is consistent with a continuous TNA model whereby evaluation data are repeatedly examined to ascertain if identified competencies have improved and if new development needs have emerged. Therefore, TNA should be a continuous process rather than a one-time diagnostic exercise for teacher evaluation.

If we apply the similar approach to a school system, the organizational level of TNA matches the school's bigger goals, for example, a school chain deciding that better assessment skills are required for the all teachers to follow a new curriculum. The task level matches the specific teaching skills a teacher's job actually needs, which is what standards documents like the National Professional Standards for Teachers in Pakistan (2009) try to describe across ten areas, including ICT use, assessment, and planning. The person level matches how well an individual teacher performs against these standards, and this is exactly the kind of data that classroom evaluation provides us with. Hence, teacher evaluation is not just similar to person-level TNA, it is actually the primary source of person-level TNA data in a well-designed system. On the other hand, the school's evaluation policy and standards document together cover the organizational and task levels.

2.4 Linking Evaluation to Training Needs Analysis

Even though evaluation and TNA fit together well in theory, most research on teacher evaluation reform has focused more on measurement problems in comparison to how the results are used for development. While reviewing evaluation reforms in America in the 2000s, (Toch & Rothman, 2008) and (Weisberg et al., 2009), both highlighted the two similar problems. First, evaluation systems could not properly tell effective teachers from ineffective ones. Second, even when they could, this data was not used properly to frame a stronger teaching workforce. (Will, 2018) focused at evaluation systems that actually produced good results and found that the small number of systems that worked well shared one common thing: they connected evaluation directly with follow-up support, instead of treating the score as the final result. This supports the main argument of this paper, that evaluation only really helps improve teaching when it is genuinely connected to training afterward.

Teacher evaluation and Training Needs Analysis are two different but complementary processes within a professional-development system. Evaluation is mainly concerned with producing evidence on the quality and effectiveness of what is currently being done. TNA interprets evidence on performance gaps in terms of the knowledge, skills, competencies or support needed to improve future performance. Robinson (2022) argues that good teacher-evaluation systems should produce feedback that is meaningful and actionable, not just summative ratings. Similarly, Nausheen et al. (2024) show how evidence from evaluation can be integrated into a faculty-development cycle where identified weaknesses are translated into targeted professional-development activities. The implication is that evaluation becomes much more useful when its findings are methodically related to the identification and prioritization of training needs.

The relationship can be viewed as a sequential but iterative process. Evaluation provides evidence of current performance. Feedback interprets the evidence. TNA identifies the underlying competency or developmental gaps. Training addresses the gaps. Subsequent evaluation determines whether improvement has occurred. Such a model moves teacher evaluation from a primarily

accountability-oriented function to a developmental function in which assessment becomes a source of evidence for professional learning.

Research on professional development provides a similar conclusion from another angle. Guidance for school districts on needs assessment emphasizes that good professional development needs proper analysis of existing PD data and school performance, and it must find and bridge the gaps between current and expected practice before designing any training (Try Our Sample Teacher Training Needs Assessment, 2022). This is almost the similar idea as the McGehee and Thayer model. Research on faculty development outside K-12 schools reaches a similar point in a different way: needs-assessment surveys alone often do not explore real gaps unless they are connected with evidence from actual classroom observation (A Novel Tool to Assess Faculty Development Needs, 2024). Research on mentoring and induction tells us more about the person level of TNA. When a training need is found through evaluation, structured mentoring in which a new or weaker teacher is paired with an experienced one through focused coaching, is one of the most reliable ways to close that gap, especially in the early career of teachers (EdResearch for Action, 2025).

Recent advances in educational technology also suggest new opportunities for strengthening the connection between classroom observation, evaluation, feedback and professional development. Qu et al. (2025) demonstrates how multimodal artificial intelligence can be leveraged to analyze classroom recordings and produce timely, personalized feedback in accordance with pedagogical practices thru the ClassMind system. These technologies could increase the amount and specificity of evidence available for teacher evaluation, and therefore provide richer data for identifying professional-development needs. However, technology may be best viewed as a support mechanism rather than a replacement for human evaluators, helping to collect and organize evidence while leaving professional judgment, contextual interpretation and developmental decision-making to educators and trained evaluators. The authors' focus on privacy and the ongoing need for human judgment is particularly pertinent when considering using AI-supported evaluation in educational environments.

2.5 The Case Study Literature

Against this background, (Asher, 2023) case study of a private chain of schools and colleges in Pakistan's North Zone provides a very comprehensive and evidence-based picture of how the relationship between evaluation and TNA actually works inside the real institution. The study used three different sources of data. First, real-time classroom observation of twenty teachers using an approved lesson-evaluation tool. Second, a thirty-respondent Likert-scale questionnaire. Third, semi-structured interviews with section heads. Using these three sources allows the findings to be evaluated from both quantitative and qualitative sides. This is a primary objective of the study, and this paper uses it a lot in the findings and discussion part below.

2.6 Research Hypotheses

Based on the case study that this paper uses, below-mentioned three hypotheses are used to guide the discussion:

1. **H1.** The teacher evaluation process explores the strengths and weaknesses of teachers, and this supports Training Needs Analysis.

2. **H2.** Constructive feedback provided during evaluation de-briefing helps teachers to enhance their teaching skills, and this supports capacity building.
3. **H3.** The teacher evaluation process positively affects and improves the quality of institutional teacher-training programs.

3. Methodology

3.1 Research Design

The paper uses descriptive-correlational case study for its empirical part which utilizes both quantitative and qualitative methods to identify the relationship between teacher evaluation and Training Needs Analysis in a private chain of schools and colleges. A correlational framework fits this study because the goal is to describe how strongly, and in what pattern, evaluation and the training needs it reveals are related rather not to prove cause and effect through an experiment. This choice matches basic guidance about choosing a research design that fits the research questions before starting data collection (Johnson & Marietta-Westberg, 2004). Deciding the research design and target population before starting fieldwork is considered in methodology literature as basic research planning, because it protects the study from losing focus and becoming too costly, which often happens when the framework is not clear from the beginning (Johnson & Marietta-Westberg, 2004).

The study had two parts that worked together. The first part used secondary data: real-time classroom observation of teachers, done by using a standard and already approved lesson-evaluation tool, provided by a teacher-training institute. The second part used primary data collection directly from teachers and section heads through semi-structured interviews and a self-administered questionnaire.

3.2 Variables

In this study, Training Needs Analysis was treated as the dependent variable, which means it was the outcome that the research wanted to explain. Further, teacher Evaluation was treated as the independent variable, meaning its changes were supposed to explain the differences explored in training needs.

3.3 Population and Sampling

The population targeted was teaching staff of the school chain in the North Zone. According to (Sekaran, 2003), population means the whole group of people, events or cases related to a research question. (Zikmund et al., 2013) give a similar definition, saying the targeted population is the complete group of the sector being studied, which here refers to the teachers of this school chain. Because of time, cost, and health-related constraints, it was not possible to collect data from the whole population. Hence, a purposive sample was chosen instead. In Islamabad, from a local teacher-training institute, thirty teachers filled the survey questionnaire, and real-time evaluation data was collected for twenty teachers. Different researchers have different opinions about the minimum sample size needed for quantitative research. Some argue thirty is an acceptable minimum, while others suggest around 250 for better generalizability (Hair et al., 2010). The sample

used in this study is on the lower side of this range, and, as the original study itself explains, it should be understood as an institutional case study rather than a survey representing the whole country.

3.4 Instruments

There were two instruments used to collect the quantitative data used in this paper. The first was a self-administered questionnaire with closed-ended questions, using a five-point Likert scale from Strongly Disagree (1) to Strongly Agree (5). It had ten statements about teaching skills matching the areas evaluated in classroom observation, followed by five items about how helpful teachers found constructive feedback, and three items about how the school used evaluation results. The second instrument was a lesson-evaluation form, filled by trained observers during real lessons. A similar five-point scale from Very Weak (1) to Excellent (5), across ten teaching areas: writing Desired Learning Objectives (DLOs), lesson planning, questioning technique, variety of teaching methods, resource material, use of ICT, time management, reflective practice, communication skills and subject knowledge. To make sure they were valid and consistent both tools were tested first and checked by senior faculty members before being used fully.

3.5 Data Collection and Analysis

Data was collected using four methods: the structured questionnaire, semi-structured interviews with section heads, some internet sources used to check background facts and the approved lesson-evaluation forms. Quantitative data from the questionnaire and evaluation forms was completely studied using simple descriptive statistics, and response percentages were calculated for each item by dividing the number of responses in each category by the total number of responses and multiplying by 100. Qualitative data from the focus-group discussions and interviews was studied using thematic analysis, and it was used to support and check the quantitative results. This matches the study's goal of thoroughly looking at the relationship between evaluation and TNA.

3.6 Reliability, Validity, and Ethics

Through pilot testing and expert review of both tools' reliability was made sure before they were used fully. All participants signed informed consent forms, were told they could leave the study at any time, and were given fake names to protect their identity, which follows normal ethical rules for research involving human participants.

4. Findings and Discussion

4.1 Strengths and Weaknesses Identified Through Evaluation (H1)

The questionnaire responses and the real-time lesson-evaluation data in this case study show a very similar pattern. The similarity between two separately collected sources is itself an important finding, because it shows that the pattern is a real feature of the institution, not just a result of one instrument.

On the classroom-observation tool, more than half of the observed teachers were rated Very Good or Excellent in using subject knowledge (95%), communication skills (80%), and instructional technology (65%). These same three areas also came out as strengths in the separate questionnaire,

where teachers agreed or strongly agreed that they were comfortable using confident in subject content (67%), technology (66.67%), and good at communication (56.67%). This matching result, between an observer's rating and the teacher's own opinion of themselves, provides insights that these are real strengths, not just something teachers believe about themselves.

The similar kind of matching pattern appears, in the opposite way, for four weaker areas. On the observation tool, more than half of the teachers scored Weak or Very Weak in writing lesson planning (70%), questioning technique (65%), Desired Learning Objectives (65%), and using different teaching methods (65%). The questionnaire results are very close to this as well: teachers disagreed or strongly disagreed that they could write objectives using SMART and ABCD criteria (66.67%), plan lessons properly stage by stage (73.33%), use of good questioning technique (60%), and use a variety of teaching methods (56.67%),

This pattern is not something unique to this one school chain. It matches a broader pattern already seen in Pakistan's teacher-development research. A large study of secondary teachers in Lahore also found that only about half of teachers used differentiated instruction, which is closely related to using different teaching methods, and that compared to the National Professional Standards for Teachers (Investigation of Teachers' Practices through National Professional Standards for Teachers in Pakistan, n.d.), instructional planning and assessment were among the weakest areas. The fact that a similar pattern of strengths and weaknesses appears now and again, in different cities and different institutions, which imply that this is a bigger, system-wide gap in how teachers are trained in Pakistan, both before and during service, especially in planning and questioning skills. This needs more thinking, not just skills used in delivering the lesson. This is important for TNA, because if the gap keeps appearing across different institutions, then writing objectives, planning, and questioning technique should be regarded as organizational-level training needs, not just individual ones, as used by (McGehee & Thayer, 1961). This implies that training investment should be system-wide, not just a few separate remedial plans for individual teachers.

In a nutshell, the data clearly supports Hypothesis 1: the evaluation process is able to reliably show which areas are strong and which need improvement, and it does this consistently across two separate tools. This provides school leaders real evidence, not just guesswork, to build a proper training plan.

4.2 Constructive Feedback and Capacity Building (H2)

Finding a gap through evaluation is important, however, it is not enough on its own. The gap must be explained to the teacher in a way they can actually use. The questionnaire data in this case study highlights directly to this point. Sixty percent of teachers agreed or strongly agreed that constructive de-briefs after lesson evaluation helped make their strengths clear, and 63.33% said the de-briefs made their weak areas clear too. Further, 83.34% of teachers, which is the highest agreement seen anywhere in this study, agreed that lesson de-briefs provided them a real chance to improve their daily teaching, and 63.33% said that after de-brief sessions, they felt able to mentor junior colleagues about what makes a lesson effective.

This finding matches the classroom-observation research mentioned earlier, which reflects that feedback needs to be specific and connected to clear standards to actually work, instead of being just a formal exercise ("ClassMind: Scaling Classroom Observation," 2025). It also matches what

Tuytens and Devos (2017) found, that evaluation only really assists development when feedback is part of a two-way leadership process, not a one-sided judgment. In this case study, it is evident that de-briefing did meet this standard well. The fact that teachers said they could use de-brief content to mentor junior colleagues reflects that the feedback was specific and useful enough to be passed on to others, not just heard and forgotten.

Similarly, there is another finding that shows the picture is not perfectly uniform: only 50% of teachers agreed that de-briefing helped them update their own records, which is the weakest result in this part of the study. This matches (Derrington & Martinez, 2019) claim that not every part of an evaluation system gives the same developmental benefit. Administrative tasks, like record-keeping, seem to benefit less from de-briefing than actual classroom teaching skills. Still, overall, the evidence supports Hypothesis 2: constructive, two-way feedback given through structured de-briefing is strongly connected to teachers feeling more able to act on their evaluation results, especially in classroom practice rather than paperwork.

4.3 Evaluation as a Driver of Institutional Training Design (H3)

The third hypothesis is about whether evaluation data actually reaches the people who make decisions, in a way that shapes the design of training, and closes the loop between evaluation and training. Here again, the case-study evidence supports this claim. Two-thirds of teachers (66.66%) agreed or strongly agreed that regular evaluation gave heads of department and section heads a chance to improve training programs. Seventy percent agreed that evaluation results helped school leaders organize model and demonstration lessons that target the weak areas found during in-service training. Most importantly, 90% of teachers agreed that colleagues with higher ratings could be used as mentors for new or junior teachers. This reveals that the person-level finding of evaluation, that is, knowing who is strong in which area, is being used as an organizational-level training resource, meaning the school knows who can teach whom.

The last finding is especially important because it closes a gap that is often mentioned as a problem in international research. (Toch & Rothman, 2008) and (Weisberg et al., 2009) both found that American evaluation systems in the 2000s often failed to turn their evaluation results into a more skilled teaching workforce, even when the evaluation tool itself was good at telling effective teachers apart from weaker ones. This case-study institution seems to be closing that same gap: teachers who perform well according to evaluation are being used as mentors for weaker or newer colleagues, which is a low-cost but high-value action. Research on induction and mentoring says this is one of the more reliable ways to help early-career teachers improve, as long as the mentoring relationship continues over time and focuses on specific skills, instead of just being a single orientation talk (EdResearch for Action, 2025). Interview data from section heads also support this. Section heads and heads of department said in the interviews that regular evaluation was important for Training Needs Analysis exactly because it explained both strengths and weaknesses in detail to make in-house training more focused, and that evaluation followed by good de-briefing tested teacher competence while also helping their professional growth at the same time. Since this matches the quantitative results too, the finding can be seen as reliable rather than a result of just one method, and overall, the evidence supports Hypothesis 3.

The fact that teacher evaluation outcomes can reveal persistent professional shortcomings lends support to the argument that evaluation should be tied to institutional training decisions. Instead of using evaluation results solely as evidence in individual performance appraisals, institutions can aggregate patterns of teacher weakness to prioritize professional-development programs. This is in line with the findings of Nausheen et al. (2024) who used evaluation evidence in their faculty-development program to identify areas that require further development and subsequently modified training activities based on feedback. This indicates that an evaluation can be used for two purposes: on the individual level, it can identify specific developmental needs; on the institutional level, aggregated evaluation data can identify common training needs.

4.4 Synthesis: The Evaluation–TNA Pipeline

Looking at the all three hypotheses together, they describe one connected process, not three separate findings. First, evaluation provides specific evidence about strengths and weaknesses (H1). Second, structured and two-way feedback turns this evidence into something teachers can actually act on (H2). Third, combined evaluation results affect school decisions about training design, model lessons, and mentoring (H3). This process matches closely with the organization-task-person idea from (McGehee & Thayer, 1961): the person-level data from individual lesson evaluations, once combined, becomes the evidence used to decide which skills need training at the task level, and this then guides organizational-level decisions about how to use training resources, such as release time, mentor assignment, and model-lesson planning, across the whole school.

However, one limitation should be noted in the case-study data. While feedback and mentoring were rated very positively, no question in the questionnaire checked whether teachers' later training was actually tracked against the specific weaknesses found in their evaluation, or whether individual development plans were written down and reviewed again in the next evaluation cycle. Without something like this, there is a risk that the evaluation-to-training process works well for a single cycle, but does not build up over time into a proper record of a teacher's growth. This matches what (Derrington & Martinez, 2019) found about evaluation systems that, despite spending a lot, did not give enough continued growth chances to teachers who were already rated as competent.

To summarize, the literature suggests an evaluation – TNA pipeline. The part of professional development where evidence is generated is teacher evaluation. Robinson (2022) stresses the need for teacher evaluation to move beyond judgment to meaningful improvement. Rucinski and Diersing (2014) discuss the evolution of evaluation systems within the framework of broader accountability and quality issues. Nausheen et al. (2024) empirically show that evaluation and feedback can inform specific professional development needs and, consequently, alter the delivery of training. More recently, Qu et al. (2025) show how multimodal artificial intelligence can scale and accelerate classroom observation and instructional feedback. Together these perspectives point to the fact that the usefulness of teacher evaluation is largely dependent on what happens after the evaluation, i.e., how evidence is turned into actionable feedback, how feedback is analyzed systematically to identify training needs, and how identified needs are addressed thru relevant professional learning. Hence, the evaluation's effectiveness should be assessed not only in terms of the accuracy of the rating it produces, but also in terms of its ability to trigger a cycle of evidence-based professional improvement.

5. Discussion

The findings in this paper, when compared with broader research, point to a bigger argument about how evaluation systems really work inside schools, and this argument is worth explaining clearly. The first point to highlight is that the pattern of strengths and weaknesses found here is not about individual teachers being good or bad. It is more about which skills a teacher-preparation system is teaching well, and which it is not. Teachers were strong in areas directly linked to delivering the lesson, such as subject knowledge, use of ICT, and communication. They were weak in areas that needed more analytical thinking before the lesson is even delivered, such as writing objectives, planning across lesson stages, asking good questions, and choosing between different methods. This is similar to an old observation in evaluation research that delivery skills are easier to pick up informally, just by classroom experience and watching senior colleagues, while questioning and planning skills need more careful, theory-based training that is less likely to be learned by accident (Toch & Rothman, 2008). If this reading is true, it matters a lot for how schools should read their evaluation results. A low score in lesson planning is less about one teacher's personal weakness, and more about a sign of a gap indicating the whole training system that evaluation has simply made visible.

Secondly, the findings about constructive feedback (H2) should be understood together with research on teacher self-efficacy, because a teacher's firm belief in their own ability is what eventually decides whether they actually act on feedback, or just listen to it politely. (Bandura, 1997) social-cognitive theory says that self-efficacy is mainly constituted through three things: succeeding at a task, watching others succeed at a similar task, and hearing encouragement from trustworthy colleagues. This description is very close to what a good de-brief actually gives a teacher: honest, specific feedback from a trusted observer, along with clear advice the teacher can apply and improve in the next lesson. A recent review and meta-analysis of studies about teacher self-efficacy found out that structured feedback-based training programs were linked to higher self-efficacy in both experienced and new teachers (Täschner et al., 2025). This argument substantiates the case-study finding that 83.34% of teachers felt that de-briefing provided them a real chance to improve daily teaching. In other terms, feedback is not just information. When it is given properly, it truly changes how much teachers believe they can improve, and this belief is probably the real reason an evaluation score turns into lasting change in behavior and approach instead of just a short, one-time adjustment.

Thirdly, this discussion has also looked at a problem that the findings do not fully solve: the difference between a long-term, lasting system and one well-run evaluation cycle. The case-study data clearly depicts what happens within one cycle, teachers are evaluated, they receive feedback, and the school/college adjusts training based on this. However, the data is unable to show if this improvement continues over time. Whether a teacher whose lesson planning that was marked as weak in one cycle actually improves in the next cycle, or the same weaknesses keep appearing now and again because training was provided but never properly applied. This is very close to the issue raised by (Derrington & Martinez, 2019), who explored that evaluation systems, even with a lot of investment, did not give yield growth chances to teachers already marked as competent. Once a teacher is taken as "good enough," evaluation can stop pushing them further, instead of continuing

to support their growth. A truly developmental evaluation-TNA system needs some way to track a teacher's progress across several cycles, not just to rightly judge a single cycle.

Fourthly, the finding that 90% of teachers supported using highly rated staff as mentors for newer or weaker colleagues should be seen as more than just a convenient staffing decision. It is actually an informal answer to the exact organizational-level problem that formal TNA models try to solve. Once evaluation provides a lot of person-level data, a school has to decide and implement how to turn this data into a useful training resource and apply practically, without starting a new outside training program. Peer mentoring is a relatively low-cost solution, and research on induction portrays it is also one of the more evidence-supported solutions. As long as the mentoring continues and focuses on a particular skill area, rather than being just one orientation session (EdResearch for Action, 2025). If we focus at this together with the McGehee and (Thayer, 1961)) model, it seems that this case-study institution is already practicing a kind of organizational-level TNA without really naming it that way, by utilizing an internal resource, mentor time, in response to person-level and task-level evaluation data. By making this an official, clearly written process, instead of leaving it as an informal process that depends on individual section heads' personal effort, is where the school has the higher chances of improvement.

Lastly, it is helpful to place these findings within the bigger question of whether teacher evaluation reform, which multiple education systems have tried over the past twenty years, has really been worth its high administrative cost. (Will, 2018) found that the small number of evaluation systems that actually got measurable results shared one common feature: of a two-part process they regarded evaluation as only the first-half and follow-up support as the second-half. The case study explored in this paper can be understood as a small example of this same idea working, not because its evaluation tool was advanced, rather the institution treated the evaluation score as the start of a conversation, not as a final judgment. This is the main principle that this paper's findings support.

All in all, these five points show that the three hypotheses tested earlier are not separate claims, but parts of one connected argument. Evaluation data is only as useful as the training system that receives it. Feedback is only as effective if it has the potential to change a teacher's belief in their own ability. And moving talent around inside the institution, through mentoring or model lessons, is only sustainable when it becomes an official policy, instead of an informal practice. The recommendations suggested in the next section should be read as an endeavor to turn these informal practices into properly written institutional procedures.

6. Implications for Policy and Practice

Several practical implications can be drawn from these findings. Firstly, the fact that the same pattern of strengths and weaknesses keeps appearing, strong in ICT use, subject knowledge, and communication, weak in writing objectives, lesson planning, using different methods, and questioning technique, suggests that in-service training in this and similar institutions would give the best results if resources are focused on planning and questioning skills, instead of spreading training thinly across all ten areas. Since a similar pattern is also found in other Pakistani research (Investigation of Teachers' Practices through National Professional Standards for Teachers in

Pakistan, n.d.), this suggestion could probably apply to teacher-education policy more widely, including how pre-service teacher training courses are designed.

Secondly, since de-briefing was strongly supported by teachers, both for making strengths clear and for allowing peer mentoring, schools should make de-briefing a structured, standards-based conversation, connected where possible to clear standards like the National Professional Standards for Teachers in Pakistan (2009), instead of leaving the quality of de-briefing to depend only on individual observers. Research on classroom observation shows that specific feedback connected to clear standards is what really makes feedback useful (ClassMind: Scaling Classroom Observation, 2025), and this gives a clear target for training the evaluators themselves.

Thirdly, since 90% of teachers supported using higher-rated teachers as mentors for newer or weaker colleagues, schools should think about making this an official process instead of an informal one. Instead of treating instructional categories only as personnel records, schools could build a structured mentoring system that pairs teachers who are strong in a certain area with colleagues who need help in that same area. This confirms what mentoring and induction research tells us: targeted, regular coaching is more effective than generic orientation programs (EdResearch for Action, 2025). Lastly, schools should also try to close the tracking gap mentioned earlier by keeping a simple individual development-plan record for each teacher, which follows them from one evaluation cycle to the next. This way, training programs can be judged not only by how well teachers received them at first, as shown here through teacher agreement in the survey, but also by whether they actually reduce the specific gap found in an earlier evaluation.

7. Limitations

This paper's main evidence comes from a single-institution case study with a fairly small sample, twenty teachers for real-time observation and thirty for the survey, which limits how much the exact percentages reported here can be generalized, even though the wider pattern of strengths and weaknesses is supported by other research done elsewhere in Pakistan. The framework used here looks at only one point in time, so it cannot show whether the training needs identified were actually fixed in later cycles. The questionnaire data, although checked against researcher-based observation, still has the usual problem of social-desirability bias, meaning teachers may answer in a way that looks better than reality. Finally, since this case study and much of the supporting Pakistani research are about private-sector, urban and semi-urban school chains, care should be taken before applying these findings to public-sector schools or rural areas, where resources and teacher qualifications may be quite different.

8. Recommendations

Based on the findings and discussion above, the following recommendations are suggested:

1. Teacher evaluation should be done regularly and in cycles, with a clear institutional aim of using it for capacity-building and Training Needs Analysis, not just for putting teachers into categories.

2. Evaluation findings should be shared clearly and openly with school principals and section heads, so that results actually shape the design of in-service training calendars, not just inform them in theory.
3. In-service training content should be planned using combined evaluation data, with special focus, in cases similar to this study, on lesson planning, writing learning objectives, using different teaching methods, and questioning technique.
4. Results from school-level evaluation-TNA cycles should be shared with relevant national or provincial curriculum authorities, so they can help design standard teacher-training materials connected to the national curriculum.
5. A structured, evaluation-based mentoring program should be made official, pairing teachers who are strong in a certain area with colleagues who need improvement in that same area.
6. Future research should try this same design in more school chains and in public-sector schools too, with a bigger sample and a longer time frame, so it can check whether training actually reduces the specific gaps found through evaluation.
7. Institutions might also consider using technology to support classroom observation and feedback. Emerging systems like ClassMind illustrate the promise of multimodal AI for analyzing classroom recordings and delivering timely, pedagogically-aligned feedback, potentially alleviating some of the time and resource burdens of traditional observation (Qu et al., 2025). But those technologies should be used to supplement human evaluation, not replace it. Implementation needs to be based on issues of privacy, contextual interpretation, transparency and professional judgment.

9. Conclusion

Teacher evaluation can serve multiple purposes together: giving feedback for better teaching, keeping a record that matters for a teacher's career, and helping with personnel decisions like reappointment or promotion. This paper has argued, and the evidence supports this, that its most important purpose, and the one most often left undeveloped in practice, is diagnostic. This means using real evidence, not just impressions, to find out where a group of teachers is strong and where they need help, and using this evidence to shape the training that a school actually provides. The case study looked at in this paper shows this process working fairly well: evaluation clearly separated strengths from weaknesses across two different tools, constructive de-briefing turned these findings into real, teacher-reported growth, and section heads used combined results to adjust training design and mentoring assignments. The fact that a similar pattern of strengths and weaknesses appears again in other Pakistani research suggests that these findings are not just something specific to one school, but a sign of a bigger professional-development priority, one that a properly planned, evaluation-based Training Needs Analysis process, based on the idea from (McGehee & Thayer, 1961) and adapted here for schools, is well placed to solve.

References

- Asher, M. F. (2023). *Teachers' evaluation: A precursor to training needs analysis (TNA) - Case study of a local chain of schools and colleges*.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
- Danielson, C. (2013). *The framework for teaching evaluation instrument*.

- Darling-Hammond, L. (2000). *Teacher quality and student achievement: A review of state policy evidence. Education Policy Analysis Archives*. 8, 1–44.
- Derrington, M. L., & Martinez, J. A. (2019). Exploring teachers' evaluation perceptions: A snapshot. *NASSP Bulletin*, 103(1), 32–50. <https://doi.org/10.1177/0192636519830770>
- EdResearch for Action, Retrieved 17 July 2026, from website: <https://edresearchforaction.org>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. Pearson Prentice Hall.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Jarocki, L. (2018). *The use of student feedback in teacher development*.
- Jewell, J. (n.d.). From inspection, supervision, and observation to value-added evaluation: A brief history of U.S. teacher performance evaluations. *Drake Law Review*, 65(2), 363–420.
- McGehee, W., & Thayer, P. W. (1961). *Training in business and industry*. John Wiley & Sons.
- National Professional Standards for Teachers in Pakistan. (2009).
- Nausheen, F., Bhupathy, R., Mohsin, H., Sheikh, S., Lyons, P., & Atapattu, D. N. (2024). A novel tool to assess faculty development needs utilizing student evaluations. *Cureus*, 16(11), e74008. doi:10.7759/cureus.74008
- ClassMind: Scaling classroom observation and instructional feedback with multimodal AI. doi:10.48550/arXiv.2509.18020
- Robinson, S. B. (2022). *Teacher evaluation: Why it matters and how we can do better*.
- Sekaran, U. (2003). *Research methods for business: A skill-building approach*. John Wiley & Sons.
- Täschner, J., Dicke, T., Reinhold, S., & Holzberger, D. (2025). “yes, I can!” A systematic review and meta-analysis of intervention studies promoting teacher self-efficacy. *Review of Educational Research*, 95(1), 3–52. doi:10.3102/00346543231221499
- Toch, T., & Rothman, R. (2008). Rush to judgment: Teacher evaluation in public education. *Education Sector Reports*.
- Tuytens, M., & Devos, G. (2017). The role of feedback from the school leader during teacher evaluation for teacher and school improvement. *Teachers and Teaching*, 23(1), 6–24.
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. The New Teacher Project*.
- Will, M. (2018). *These six teacher-evaluation systems have gotten results, analysis says*.
- Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2013). Cengage Learning. In *Business research methods*.