# Integrating Wholes and Parts in Teacher Professional Development: Evidence from an Experimental Study with Secondary Mathematics Teachers in Ethiopia

Hailu Nigus Weldeana
Mekelle University
hailu.nigus@mu.edu.et

Frederick K. S. Leung
The University of Hong Kong,
frederickleung@hku.hk

Teacher professional development (TPD) plays a crucial role in enhancing teachers' knowledge and practice, thereby improving teaching effectiveness. This study's data analysis revealed that TPD interventions that treat teacher knowledge domains as an integrated whole significantly benefit secondary mathematics teachers compared to those that approach knowledge in a fragmented manner or provide no intervention at all. Among the groups studied (i.e., Integrated, Segregated, and control groups), all showed improvements over time. However, the Integrated Group, which experienced a holistic TPD approach, demonstrated substantial gains in knowledge compared to the Segregated Group, which focused solely on content knowledge. The Segregated Group, in turn, performed better than the Control Group, which received no intervention. Even when controlling for variables such as experience, education, gender, and grades taught, the Integrated Group consistently outperformed its Segregated and control counterparts, reinforcing the value of a comprehensive approach to TPD.

***Keywords***: Changes in Knowledge; Changes in Practice; Teacher Knowledge Domains; Student Outcomes; TPD.

## Introduction

In response to persistent educational challenges, Ethiopia launched the Education and Training Policy in 1994, emphasizing student-centered, active, and problem-based learning (Federal Democratic Republic of Ethiopia [FDRE], 1994), later updated to reflect contemporary priorities (FDRE, 2023). Building on this policy, the country implemented major reform initiatives aimed at improving teacher education and instructional quality, including the Teacher Education System Overhaul (TESO) (Ministry of Education [MoE], 2003), the General Education Quality Improvement Program (GEQIP II, 2013), the Teacher Professional Development (TPD) framework and Practical Toolkit (MoE, 2009a, 2009b), and successive Education Sector Development Programs (ESDPs) (MoE, 2015a, 2021).

Despite expanded access and rising enrollment (UNESCO, 2021; UNICEF, 2020), student learning outcomes, particularly in secondary mathematics, remain low (MoE, 2018; Dawit et al., 2022). Scholars attribute this stagnation to fragmented professional development that overlooks teachers' cultural and contextual realities (Woldehanna & Jones, 2019; Kassaye, 2006), limited subject matter expertise, weak pedagogical skills (MoE, 2018; JICA, 2012; Tefera & Asgedom, 2020), and low motivation or commitment (Mekonnen, 2019; Yismaw, 2021).

To address these gaps, this study introduces an integrated TPD model for secondary mathematics teachers, grounded in three principles: integrating multiple knowledge domains, deepening existing knowledge, and engaging in cognitively demanding tasks. Teachers participated in scenario-based problem-solving activities that linked content knowledge (CK), pedagogical knowledge (PK), pedagogical content knowledge (PCK), and knowledge of contextual features (KCF).

Focusing on secondary mathematics in Tigray, Ethiopia, is timely and critical, as this level plays a pivotal role in national development by fostering a skilled workforce, promoting educational continuity, instilling civic values, and advancing equitable learning opportunities (MoE, 2018; Dawit et al., 2022; UNESCO, 2021).

The study involved three teacher groups: Integrated TPD, content-focused training, and a control group. It aimed to evaluate the integrated model's effect on teacher knowledge, instructional practices, and student learning outcomes, arguing that a holistic approach, addressing interconnected knowledge domains rather than isolated competencies, can enhance professional growth and student achievement.

## Problem Formulation

Persistent weaknesses in teacher professional development (TPD) remain a challenge, particularly in Ethiopia's Tigray region, where programs have failed to produce measurable gains in teacher knowledge or student mathematics achievement (Weldeana, 2020). Globally, TPD initiatives often show minimal impact due to fragmented design, misalignment with teacher needs, weak theoretical grounding, and lack of integration across knowledge domains (Gemeda & Tynjälä, 2015; Taddese & Rao, 2023). Within Ethiopia, evaluations reveal persistent deficits in teacher competence, encompassing subject expertise, pedagogical skill, and professional attitudes (MoE, 2015b, 2018, 2021; Japan International Cooperation Agency [JICA], 2012), a trend echoed worldwide (Popova et al., 2022).

Despite substantial investment in national reforms, including continuous TPD frameworks and sector-wide initiatives (Betemariam, 2017), inefficiencies persist, reflecting global trends where funding often fails to translate into improved practice or outcomes (Desimone, 2009; Merriam & Leahy, 2005). Only a minority of large-scale TPD programs met professional standards, and few adequately address secondary education (Guskey & Yoon, 2009; Yoon et al., 2007).

To address these gaps, this study implemented a context-sensitive, integrated mathematics TPD intervention in Mekelle City, Tigray. The model incorporates four knowledge dimensions: content, pedagogy, pedagogical content knowledge, and contextual knowledge, through problem-based, scenario-driven activities designed to promote cohesive professional growth. The study investigates the following research questions: (a) Does the TPD intervention enhance teachers' knowledge? (b) How does it influence instructional practices? (c) Do teachers' participation affect students' learning outcomes?

*Hypothesis*: The study hypothesizes that the integrated TPD model will improve teachers' knowledge across domains, expand instructional capabilities, and enhance student mathematics performance. Over the past five decades, TPD has evolved as a central strategy for strengthening teaching effectiveness, deepening professional knowledge, and improving student learning (Jones & O'Brien, 2024). Research underscores its role in shaping teachers' beliefs and practices, which directly affect student achievement (AbdulRab, 2023; Audisio et al., 2023). Meta-analyses further validate these claims, highlighting consistent features of effective TPD (Sims et al., 2025; Popova et al., 2022; Darling-Hammond et al., 2017a, 2017b; Ehlert & Souvignier, 2024).

The relevance of TPD is context-and discipline-dependent. For example, science teachers who engage in content-ICT-integrated TPD show measurable student learning gains (Zhou et al., 2024). Similarly, literature reviews identify key components of effective TPD, especially those involving digital tools (Amemasor et al., 2025). In 2024, Ethiopia's Ministry of Education (MoE, 2024) launched a nationwide capacity-building program targeting secondary teachers, emphasizing integrated subject knowledge, pedagogy, and digital tools to enhance instructional practices and student outcomes.

This study distinguishes itself by defining specific domains of teacher knowledge, examining their interaction during TPD, and exploring their application within the complexities of classroom teaching, providing a comprehensive approach to professional growth and improved student achievement.

## Related Literature

### *Theoretical Framework*

Shulman's (1986, 1987) seminal work established content knowledge (CK), pedagogical knowledge (PK), and pedagogical content knowledge (PCK) as essential domains for effective teaching and meaningful student learning. Subsequent studies expanded these domains to include contextual and curricular dimensions (Grossman, 1990; Clarke & Hollingsworth, 2002; König & Kramer, 2016). Hill et al.'s (2008) Mathematical Knowledge for Teaching (MKT) framework further disaggregates teacher knowledge into common content knowledge (CCK), specialized content knowledge (SCK), knowledge of content and students (KCS), and knowledge of content and teaching (KCT), emphasizing the complexity of mathematics instruction.

Teacher professional development (TPD) is widely recognized as a key mechanism for enhancing teacher knowledge, instructional practice, and student achievement (Kennedy, 2016; Darling-Hammond et al., 2017a, 2017b). Effective TPD is sustained, reflective, and context-responsive (Avalos, 2019; OECD, 2021). However, traditional approaches—often fragmented, standardized, and top-down—neglect teachers' diverse competencies and classroom realities, limiting effectiveness (AbdulRab, 2023; Opfer & Pedder, 2011).

In Ethiopia, TPD initiatives have improved formal qualifications and introduced active learning (MoE, 2018), yet student performance remains low due to limited pedagogical capacity, inadequate teacher motivation, and systemic constraints such as top-down implementation (MoE, 2015; Tessema, 2023). Scholars call for integrated, context-sensitive TPD models that develop teachers cognitively, pedagogically, and contextually (Woldehanna & Jones, 2019; Tesfaw & Hofman, 2020).

This study adopts a four-dimensional framework of teacher knowledge (i.e., CK, PK, PCK, and knowledge of contextual features (KCF)), emphasizing their interdependence for designing effective, context-responsive TPD. Teachers engage in cognitively demanding, real-world tasks and reflective practices to enhance instructional expertise and student outcomes.

Knowledge Integration highlights the need to connect multiple knowledge domains during professional learning. Teachers exposed to integrated TPD demonstrate improved instructional quality and student learning compared to conventional models (Ball et al., 2008; Weldeana, 2020). High-impact TPD is coherent, content-focused, and embedded in daily practice (Carpenter et al., 1996; Darling-Hammond et al., 2017a, 2017b).

Knowledge Augmentation emphasizes culturally responsive education. Incorporating students' cultural and community knowledge enhances engagement and learning (Aguirre et al., 2021; Nasir & Vakil, 2020), aligning with ethnomathematics research (Gerdes, 1994; D'Ambrosio, 2001, 2006) and reinforcing the importance of cultural relevance in mathematics instruction.

Cognitively Demanding Tasks require teachers to apply knowledge across domains through complex, open-ended problem-solving, fostering deep mathematical reasoning and pedagogical reflection (Tekkumru-Kisa & Stein, 2015; Hiebert & Morris, 2022). When scaffolded appropriately, such tasks promote teacher growth without inducing burnout (Chapman, 2013; Kazemi et al., 2021).

Together, these principles form a cohesive, adaptive framework that emphasizes depth, relevance, and cultural grounding, aligned with global recommendations for effective TPD in developing contexts such as Tigray, Ethiopia, where localized content and teacher agency are essential for sustained instructional improvement (World Bank, 2021; OECD, 2020).

## *Research Evidence on the Suitability of the Model*

### *Models that Initiate Knowledge Integration*

Contemporary cognitive science and educational psychology emphasize integrating multiple teacher knowledge domains to enhance instructional effectiveness. Bloom's Taxonomy (Bloom, 1956) conceptualized competence across cognitive, affective, and psychomotor domains and has been widely applied in education. Wilson et al. (1968, as cited in Brumbaugh et al., 1997) adapted it for mathematics, aligning problems with cognitive levels, e.g., "Given $log2 = a$, what is $log25$ ?" integrates content knowledge (CK) and pedagogical knowledge (PK), including scaffolding strategies and student psychology. Anderson and Krathwohl's (2001) Revised Bloom's Taxonomy further emphasizes higher-order tasks requiring problem-posing and problem-solving aligned with analysis or evaluation.

Van Hiele's theory of geometric thinking (van Hiele, 1986) outlines hierarchical cognitive levels in geometry, with learning dependent on structured instructional progression. Misaligned tasks can impede conceptual growth, requiring teachers to integrate CK, PK, and contextual knowledge to scaffold student reasoning effectively (Battista, 2022). Cultural context further shapes geometric understanding; learners' engagement with spatial reasoning through art, architecture, or traditional crafts underscores the need for culturally responsive instruction (Shulman, 1986, 1987). Pedagogical content knowledge (PCK) unifies these domains, enabling teachers to translate abstract content into meaningful, contextually grounded instruction.

Recent studies reinforce these principles. Lai and Hwang (2023) highlight that integrating CK, PK, and contextual knowledge is essential for designing cognitively engaging tasks, particularly in digital and blended learning environments, and for aligning theoretical frameworks with classroom realities.

### *Analysis and Synthesis of Knowledge Domains in TPD*

Evidence supporting explicit integration of teacher knowledge domains in TPD is limited but growing. Dewey (1978) argued that analysis organizes knowledge, while synthesis unifies it, enhancing practical application. TPD programs integrating CK, PK, PCK, and knowledge of contextual features (KCF) are more likely to drive meaningful instructional change.

Kennedy (1998) and Bransford et al. (1999) emphasized that effective mathematics and science TPD combines content mastery, pedagogy, and student cognition. Heller et al. (2012) confirmed this with data from 270 teachers and 7,000 students, showing integrated TPD improves instructional effectiveness. Garet et al. (2001) reported that TPD aligned with adult learning principles, active engagement, and coherent standards produced significant gains in CK and teaching skills, demonstrating that integration outperforms isolated approaches.

Models such as LINCS (Ball et al., 2001; Boston, 2006) and van Hiele-based interventions show that structured progression in geometric reasoning enhances teacher cognition and instructional capacity. Similarly, Cobb et al. (1991) found that TPD rooted in integrated knowledge frameworks improves student learning. Longstanding programs like IMPACT (Fennema et al., 1989; Carey et al., 1995) and QUASAR (Silver & Stein, 1996; Stein et al., 1999) emphasize cognitively guided instruction and reflective practice, using open-ended tasks to foster higher-order thinking and discourse.

The success of these initiatives aligns with constructivist principles (Molnar, 2002) and the belief in universal student potential (Rist, 2007). Desimone and Garet (2015) highlight that effective TPD is content-focused, active, coherent, and collaborative, while Opfer and Pedder (2019) advocate adaptive, context-sensitive models that integrate knowledge domains for sustainable teacher learning. Australian programs such as ARTISM, EMIC, and the Negotiation of Meaning Project (Clarke & Hollingsworth, 2002) similarly underscore authentic tasks, reasoning, and communication in TPD. These frameworks informed the design of the present integrated TPD model, which merges CK, PK, PCK, and KCF to enhance teacher learning and instructional innovation.

## Research Methodology

### *School Selection and Group Assignment*

Of 13 public secondary schools in Mekelle, Ethiopia, 11 were included in the intervention; one was used for piloting, and one in a correctional institution was excluded. All mathematics teachers from these schools and Grade 10 and 12 students formed the target population (Creswell & Creswell, 2018).

Schools were stratified by type, teacher demographics, and student enrollment, then randomly assigned to three groups—Control, Integrated, and segregated—to ensure internal validity and reduce potential confounding (Fraenkel, Wallen, & Hyun, 2019). Randomization at the school level minimized cross-arm contamination, while students were randomly sampled within schools for outcome measurement (Shadish, Cook, & Campbell, 2002; Gay, Mills, & Airasian, 2012; MoE, 2024).

### *Research Participants and Supporting Personnel*

School leaders, including directors, vice directors, and mathematics department heads ($n = 33$), facilitated implementation, organized logistics, and supported data collection, aligning with Guskey's (2002) recommendations.

*TPD Facilitator*: The program was led by the principal researcher, a mathematics educator with a B.Sc., M.Ed., and Ph.D., over two decades in teacher education, four years in secondary teaching, and extensive leadership experience in professional development and quality assurance.

*Role During Data Collection*: The facilitator coordinated exam schedules to minimize disruption (Cohen, Manion, & Morrison, 2007). Observations were conducted during actual lessons, and weekend assessments optimized participation.

### *Teacher Participants*

Ninety-nine mathematics teachers participated in the pretest: Control ($n = 33$), Integrated ($n = 36$), and Segregated ($n = 30$). Complete posttest data were obtained from 92 teachers: Control ($n = 32$), Integrated ($n = 31$), and Segregated ($n = 29$). Attrition from pretest to posttest was minimal (7.1%) and evenly distributed across groups, $\chi^2(2, N = 99) = 0.28$, $p = .87$, indicating no significant differential attrition.

Fifteen teachers (five per group) with at least 10 years of teaching experience and consent participated in three structured classroom observations. The facilitator maintained neutral relations with all participants; prior familiarity with teachers was evenly distributed across treatment conditions. Group assignments were undisclosed to both participants and the facilitator to minimize expectancy effects (Denscombe, 2014).

### *Student Participants*

A multi-stage probability sampling method was used to target the population of 7,256 students. The sample size was determined using Cohen et al. (2007) for 95% confidence and a 5% margin of error, producing a target sample of 367 students. We have applied the formula:

$$SZ = \frac{N \cdot z^2 \cdot p(1-p)}{e^2 N + z^2 \cdot p(1-P)}$$

Where:

$SZ$ = required sample Size
$N$ = 7256 (Population Size)
$z$ = 1.96 (95% confidence level)
$p$ = 0.5 (maximum variability)
$e$ = 0.05 (margin of error)

Due to consent constraints, 300 students were evenly distributed across the treatment groups at pretest: Control (n = 100), Integrated (n = 100), and Segregated (n = 100). At posttest, 283 students provided complete data: Control (n = 95), Integrated (n = 95), and Segregated (n = 93). Overall attrition was 5.7%, and group-wise attrition was not statistically significant, $\chi^2(2, N = 300) = 0.21$, $p = .90$, suggesting no systematic loss by treatment condition. CONSORT-style participant flow was as follows:

- Assessed for eligibility: 7,256 students
- Randomly selected: 367
- Consented and enrolled: 300 (100 per group)
- Completed posttest: 283 (95, 95, and 93, respectively)
- Excluded from analysis: 17 (withdrawal, absence, or incomplete responses).

Given the modest attrition rate, potential bias was assessed using inverse probability weighting (IPW) and multiple imputation (MI) sensitivity analyses. Both methods yielded substantively similar ANCOVA results, reinforcing the robustness of the findings despite minimal missing data.

### *Nature of Intervention, Duration, and Structure*

#### *Content*

The Segregated Group received a content-focused intervention emphasizing textbook-based problem-solving to refresh subject knowledge using conventional methods. The Integrated Group underwent a knowledge-integrative program designed to establish coherence among teacher knowledge domains (CK, PK, PCK, KCF), centered on core topics spanning Grades 9–12. Both interventions were guided by a training handbook developed from established resources (Rosen, 2006; Huetinck & Munshin, 2000; van de Walle, 1998; Dolan et al., 1997; Heddens & Speer, 1995a, 1995b; Martin, 1995; Sobel & Maletsky, 1988; Fennell et al., 1987; Gerland, 1987; Wells, 1986). The handbook for the Integrated Group was organized into three sections: (a) Preliminary, General Pedagogy; (b) Main, Knowledge Integration; and (c) Test Construction, Measurement, and Evaluation. However, the Segregated group approach focused solely on content. Sections (a) and (c) supported the main section, promoting flexible, self-directed learning and reflective practice. Sessions were structured around SMART objectives (MoE, 2006).

The main section included 21 sessions (see Appendix A), supplemented by the preliminary session and 10 sets of sample tests to reinforce assessment literacy and task design. Teachers received print and digital manuals, facilitating in-session, school-level, and independent engagement.

*Training*

Guided by the Theory of Change (Fullan, 2006) and Theory of Practice (Leinhardt et al., 1995), Integrated Group teachers engaged in cognitively demanding, reflective tasks via a five-point inquiry process:

1. Content Inquiry: Defining concepts and prerequisites.
2. Pedagogy Inquiry: Exploring lesson structure and theoretical grounding.
3. Instructional Techniques Inquiry: Bridging theory and practice.
4. Context and Culture Inquiry: Integrating local and cultural relevance.
5. Review: Ensuring coherence and refining strategies.

Instruction followed a pyramiding approach (think-pair-share) consistent with adult learning principles (Knowles, 1980), promoting active, culturally responsive learning. Weekend intervention sessions for the Integrated Group were held weekly over approximately 10 months (≈40 weeks). Each session lasted 150–180 minutes (2.5–3 hours), totaling about 120 hours of centralized contact time. Hence, centralized total hours = 40 sessions×2.5–3.0 hours/session = 100–120 hours.

In addition to these centrally organized sessions, teachers received school-level mentoring and support, which included collaborative lesson planning, observation–feedback cycles, and reflective practice activities. These occurred approximately once per week for an additional 1.5–2 hours, bringing the total engagement time to approximately 160–200 hours. Thus, school-site total hours = 40 weeks×1.5–2.0 hours/week = 60–80 hours.

This gives the total intervention exposure = 100–120 (centralized) + 60–80 (school-site) = 160–200 hours (≈160–200 hours average)

Thus, the total intervention dosage ranged from 160 to 200 hours, combining structured weekend sessions and continuous school-based professional learning activities.

The Segregated Group received traditional, lecture-based instruction with minimal interaction, focusing solely on content. Control Group teachers received no intervention.

## *Instruments of Data Collection*

Assessment drew on validated teacher knowledge frameworks (Shulman, 1987; Clarke & Hollingsworth, 2002; Gess-Newsome, 1999; Hill et al., 2004, 2005) and instruments (Hill & Ball, 2004; Hill et al., 2004, 2008; Kleickmann et al., 2012). Items were adapted from mathematics texts (Rosen, 2006; Khanna & Sharma, 1998; Bali, 1997; Sobel & Maletsky, 1988; Ellis & Gulick, 1986) and assessed four domains: algebra, geometry, context and culture, and calculus.

*Teacher Mathematical Knowledge Test (TMKT)*

The TMKT included 44 items: 19 in algebra (81 sub-items), nine in geometry (44 sub-items), nine in calculus (45 sub-items), and seven cultural/contextual items (25 sub-items). Cultural/contextual problems developed in consultation with cultural sources (Gerdes, 1988, 1994a, 1994b, 2014; Zaslavsky, 1991a, 1991b; Barton, 1996; Bishop, 1988). The TMKT was administered pre- and post-intervention to evaluate the effect of TPD on teacher knowledge integration and instructional competence.

The closed-ended section of the Teacher Mathematical Knowledge Test (TMKT) employed a three-point Likert-type scale: disagree, undecided, or agree. Items were framed around mathematical scenarios attempted by a third party, typically a student, requiring teachers to evaluate reasoning rather than solve problems directly and critically. This design shifted focus from answer verification to analysis of underlying logic.

The internal consistency estimates for the TMKT subscales (Cronbach's $\alpha = 0.66$–$0.79$) are within the acceptable range for exploratory investigations of teachers' mathematical knowledge. According to Nunnally and Bernstein (1994), $\alpha$ values above 0.60 are appropriate for early-stage or exploratory research, particularly when constructs are multifaceted. The corresponding McDonald's $\omega$ coefficients (0.68–0.81) corroborate these results, providing more robust estimates of reliability when the assumption of tau-equivalence is relaxed (McDonald, 1999; Dunn et al., 2014).

Confidence intervals around the $\alpha$ (95% CI) value indicate satisfactory stability of internal consistency across domains. Algebra ($\alpha = 0.79$; $\omega = 0.81$) exhibited the highest reliability, whereas Geometry ($\alpha = 0.66$; $\omega = 0.68$) and Calculus ($\alpha = 0.66$; $\omega = 0.69$) showed moderate yet acceptable levels. These variations likely stem from item heterogeneity and differences in teachers' prior exposure and familiarity with each mathematical domain.

From a validity standpoint, the pattern of reliability across domains supports the intended construct differentiation of the TMKT instrument. The stronger reliability in Algebra reflects coherent item performance within this domain, while moderate reliability in Geometry and Calculus suggests broader conceptual diversity. The Culture and Context subscale ($\alpha = 0.69$; $\omega = 0.71$) displayed moderate internal consistency, indicating a stable yet evolving measure of teachers' sociocultural understanding of mathematics.

Overall, the TMKT subscales demonstrate adequate reliability and emerging construct validity for exploratory research on secondary teachers' mathematical knowledge. Further refinement of items—particularly in the Geometry and Calculus domains—and analysis at the subtest or facet level could improve precision and strengthen validity for future applications linking TMKT to instructional quality and student learning outcomes (See Appendix B).

Scoring for closed-ended sub-items awarded 3 points (marked with +) for agreement with correct responses or disagreement with incorrect responses, including justified reasoning. Disagreement with correct answers or agreement with incorrect answers to a sub-item received 1 point (marked with –), while neutral responses scored 2 (marked with 0), 3 otherwise (marked with 3) and the possible range of scores for algebra, for example, including 19 items (n = 81sub-itms) is (n = 81sub-itms) × 1–3 = 81–243 points, applied to all domains without exception. Each item included an open-ended component in which teachers proposed instructional strategies, explained methods, and demonstrated subject-matter knowledge and cultural awareness. Open-ended responses were evaluated using a five-point rubric: 0 = incompetence; 1 = novice; 2 = average; 3 = basic competence; 4 = proficient, yielding scores of 0–4 per item, as displayed in Table 1.

Sample items illustrated the instrument's application. One item required teachers to evaluate a student's labeling of exponential function graphs. At the same time, another involved assessing conjectures about an equilateral triangle inscribed in one circle and circumscribing another, linking the triangle's side lengths to the circle radii. Additional items assessed cultural/contextual reasoning and calculus knowledge, integrating content, pedagogy, and contextual awareness in a mixed-method, post-positivist framework (see Appendix B).

*Teacher Written Reflections on Lesson Presentations*

Lesson observation is a reliable tool for evaluating instructional practice and minimizing biases inherent in self-reporting (Cohen et al., 2007; Guskey, 2002, 2012; Kennedy, 1999). In this study, lesson observations collected real-time data on teachers' instructional approaches, with continuous monitoring across approximately three 3-month phases. Teacher-written reflections focused on student outcomes, instructional strategies, classroom activities, grouping, and engagement.

For the Integrated Group, reflections were extended to peer presentations with facilitator feedback, enhancing attribution of observed changes to the intervention. All observations were corroborated through audio and video recordings to ensure accuracy and completeness.

Table 1
Scoring Rubric for TMKT Open-Ended Questions

| SN | Teachers' Competence Level | Indicators of teacher competence: The Teacher: |
|---|---|---|
| 1 | ***Proficient*** **(4)** | • *Extracts precise and comprehensive information from the problem context.*<br>• *Demonstrates effective use of data to support conclusions.*<br>• *Proposes well-aligned strategies relevant to the problem.*<br>• *Showcases innovative thinking in addressing the problem.*<br>• *Achieves clarity and precision in describing strategies.* |
| 2 | ***Basic Competence*** **(3)** | • *Extracts sufficient and accurate information from the problem context.*<br>• *Demonstrates adequate use of data to draw conclusions.*<br>• *Provides a satisfactory explanation of proposed strategies.*<br>• *Frequently evaluates strategies within the problem context.*<br>• *Commits minimal errors in describing strategies.* |
| 3 | ***Average Competence*** **(2)** | • *Infrequently extracts information from the problem context.*<br>• *Uses information inaccurately to support conclusions.*<br>• *Proposes inconsistent strategies for the problem.*<br>• *Produces only a partial solution.*<br>• *Makes numerous writing errors.* |
| 4 | ***Novice*** **(1)** | • *Extracts irrelevant information from the problem context.*<br>• *Rarely uses data to draw conclusions.*<br>• *Proposes inappropriate strategies.*<br>• *Suggests solutions vaguely related to the problem.*<br>• *Frequently makes writing errors.* |
| 5 | ***Incompetence*** **(0)** | • *Extracts completely inaccurate information from the problem context.*<br>• *Fails to use data for drawing conclusions.*<br>• *Attempts irrelevant or misleading ideas.*<br>• *Offers no solution, or the solution is unrelated to the problem.* |

• *Writing is error-ridden and unclear.*

*Student Mathematical Achievement Test (SMAT)*

SMAT assessed the mathematical achievement of Grades 10 and 12 students pre- and post-intervention. The Grade 10 version comprised 60 items (algebra, geometry, cultural mathematics), while the Grade 12 version comprised 75 items, including calculus (administered during the posttest). Items were derived from national exams, official textbooks, and problem-rich resources (Khanna & Sharma, 1998; Bali, 1997; Sobel & Maletsky, 1988; Coxford et al., 1991; Zameeruddin et al., 1980). Scoring used a binary system (correct = 1, incorrect = 0). Pilot testing confirmed acceptable reliability ($\alpha = 0.99$ for grade 10 and $\alpha = 0.74$ for grade 12), and validity aligned with established mathematics assessment frameworks.

*Procedure*

To evaluate the intervention's effects, 11 secondary schools were randomly assigned to one of three conditions: the integrated, segregated, or control group. Randomization was implemented at the school level rather than by individual teachers to avoid potential contamination (i.e., the exchange of information or influence) that could arise if teachers from the same school were allocated across different experimental groups. This design choice aligns with best practices in educational intervention research to maintain the study's internal validity (Creswell & Guetterman, 2019; Shadish, Cook, & Campbell, 2002).

Each group received a distinct form of treatment. Teachers in the integrated group participated in sessions that deliberately fused multiple domains of professional knowledge, including CK, PK, PCK, and KCF. These teachers participated in both pre-intervention and post-intervention testing. The segregated group, on the other hand, underwent a content-centered intervention aimed primarily at strengthening subject matter knowledge (CK) without explicitly linking it to pedagogy or student understanding/Thinking. This group also completed pretests and posttests. The control group received no intervention but served as a benchmark through their participation in pre- and post-testing (Cook & Campbell, 1979; Creswell, 2014).

Ethical protocols were rigorously observed throughout the study. Informed consent was secured from all participating stakeholders, including students, teachers, school administrators, and local education authorities. Formal approvals were obtained not only from the individual schools but also from the relevant regional and district-level education bureaus (locally known as Woredas or sub-cities). These measures ensured compliance with ethical standards for research involving human participants (British Educational Research Association [BERA], 2018).

*Preference Judgments Variable*

The construct "preference judgment" reflects teachers' ability to identify correct mathematical thinking when evaluating a third party's (primarily a student's) work presented in classroom scenarios. Teachers rated each scenario using a three-point Likert scale: Disagree (1), Undecided (2), and Agree (3). This enables them to grasp their content along domains of algebra, geometry, calculus, as well as culture and context.

Item Coding involved the following: Each scenario was treated as a separate item, scores were summed across all items to form a composite preference-judgment score, and the overall reliability of the scale was assessed using Cronbach's α, yielding acceptable internal consistency for exploratory measurement.

Although the original data are ordinal, aggregated summed scores across multiple items can approximate interval-level measurement (Norman, 2010), justifying the use of OLS regression and ANCOVA for preference judgments.

To ensure results are not sensitive to this approximation, Ordinal/logit models were fit at the item level using generalized estimating equations (GEE), which account for clustering within teachers, and non-parametric tests (Kruskal–Wallis and Wilcoxon rank-sum) confirmed the observed group differences were consistent with the OLS results.

These checks support the validity of treating aggregated preference-judgment scores as quasi-interval data for linear modeling while transparently acknowledging their ordinal nature.

## *Data Analyses*

### *Quantitative Data Analyses*

Effectiveness of the intervention was assessed using Ordinary Least Squares (OLS) regression on pre- and posttest scores from teachers' TMKT and students' SMAT. Independent variables included intervention type (Integrated, Segregated, Control), while dependent variables were teacher and student outcomes. Control variables comprised gender, teaching experience, qualifications, and grade level taught. Pretest analyses established baseline equivalence across groups (Creswell & Guetterman, 2019). Posttest OLS analyses estimated treatment effects, reporting regression coefficients (β), standard errors (σ), t-values, ANCOVA, effect size (η), and significance levels ($p \leq 0.05$), allowing causal inference while controlling for confounds (Wooldridge, 2013). All analyses were performed in SPSS, integrating multiple data sources to comprehensively evaluate intervention impact, consistent with mixed-methods approaches (Creswell & Guetterman, 2019).

### *Qualitative Data Analyses*

Teachers' written reflections on lessons and intervention sessions were analyzed thematically following Braun and Clarke's (2006) six-phase framework. A basic interpretive qualitative approach (Cohen et al., 2007) guided inductive coding, using both a priori codes informed by Parsons (2011) and emergent themes. The process included familiarization, coding, theme development, refinement, and reporting. Written reflections were systematically categorized by intervention phase and group to maintain context and coherence. A five-step content analysis procedure (Cohen et al., 2007) was employed: extracting annotations, organizing into categories, recording prominence, clustering themes, and analyzing meaning. This approach ensured a rigorous, contextually grounded interpretation of teacher experiences.

## *Statistical Analysis and Robustness Checks*

Analyses of covariance (ANCOVA) were conducted to examine the effects of treatment type on teachers' posttest scores across the domains of Teacher Mathematical Knowledge for Teaching (TMKT), controlling for pretest performance and relevant covariates. However, given the modest sample size and the nesting of teachers within schools, small-sample corrections were applied to improve inference reliability.

To address potential bias in standard errors and test statistics due to clustering, cluster-robust variance estimators (CR2 and CR3) were computed using the *clubSandwich* approach (Pustejovsky & Tipton, 2022). These adjustments yield small-sample-corrected F-tests and confidence intervals that are less sensitive to the limited number of school clusters.

Additionally, permutation-based randomization tests were conducted at the school level, following recommendations by Manly and Navarro (2015). This nonparametric procedure randomly reassigns treatment labels while preserving intra-school dependencies, providing empirical *p*-values that do not rely on large-sample approximations.

Results from both robustness checks confirmed the stability and consistency of the main ANCOVA findings presented in Appendix C. The treatment effect remained statistically significant across all TMKT domains, though *p*-values for smaller domains (e.g., Culture and Geometry) were modestly attenuated under CR2/CR3 adjustments. The integrated professional development model consistently demonstrated higher adjusted posttest means than both segregated and control groups, affirming the robustness of the observed effects.

Nonetheless, due to the limited number of school clusters, the statistical power for detecting smaller effects may be constrained. Hence, results should be interpreted as indicative yet robust evidence of the positive impact of integrated, collaborative teacher professional development. The combined use of CR2/CR3 corrections and permutation tests enhances confidence in the findings while maintaining transparency about the limits of inference in small-sample clustered designs.

## Results and Findings

### *Baseline Assessment of Teachers and Students: Insights from Pretests*

#### *Baseline Evaluation of Teachers: Insights from TMKT Pretest Results*

Table 2 summarizes baseline TMKT and control variables across groups. No significant differences were found in total TMKT scores ($p = .331$), indicating overall comparability. However, imbalances emerged: the Control group had a higher proportion of male teachers (97%) compared to the Integrated (75%) and Segregated (70%) groups ($p = .013$). Integrated group teachers reported more years of experience ($M = 3.22$; typically 10–25 years) than the Control ($M = 2.88$; 5–10 years, trending toward 10–25) and Segregated ($M = 3.03$; 10–25 years) groups ($p = .037$). Gender was coded nominal, while the rest were ordinal. In content domains, only Geometry showed a significant difference, with the Segregated group outperforming the others ($M = 88.20$, $SD = 7.62$; $p = .020$). Algebra, Calculus, and Cultural Context scores were statistically comparable across groups.

Although group differences in Algebra approached significance (Integrated: $M = 162.80$; Control: $M = 155.90$; Segregated: $M = 162.10$; $p = .104$), no significant differences emerged in other TMKT domains or covariates. Given significant baseline imbalances in gender, teaching experience, and Geometry, these variables were included as covariates in post-intervention regressions to strengthen the robustness of treatment effect estimates, with Geometry emphasized as a key outcome.

Table 2

Baseline/Pretest TMKT Results: Mean (SD) and p-Values for Treatment Groups

| Variables | Control (N=33) | Integrated (N=36) | Segregated (N=30) | *P* |
|---|---|---|---|---|
| Gender | 0.97 (0.17) | 0.75 (0.44) | 0.70 (0.47) | 0.013* |
| Experience | 2.88 (0.60) | 3.22 (0.49) | 3.03 (0.56) | 0.037* |
| Educational Level | 2.15 (0.51) | 2.19 (0.62) | 2.28 (0.58) | 0.727 |
| Grades Taught | 2.82 (1.61) | 3.44 (1.99) | 2.90 (1.54) | 0.271 |
| Algebra | 155.90 (15.99) | 162.80 (15.45) | 162.10 (10.55) | 0.104 |
| Geometry | 78.24 (19.00) | 78.69 (17.06) | 88.20 (7.62) | 0.020* |
| Calculus | 79.18 (15.84) | 79.64 (16.44) | 77.10 (17.24) | 0.808 |
| Cultural Context | 35.30 (13.53) | 35.17 (13.34) | 37.07 (13.64) | 0.824 |
| Total | 356.50 (45.53) | 349.00 (42.77) | 364.5 (36.42) | 0.331 |

* Statistically significant at $p < 0.05$.

### *Baseline Assessment of Students: Analysis of SMAT Pretest Results*

Baseline comparability of student cohorts was assessed using SMAT pretest scores for Grades 10 and 12 across Control, Integrated, and Segregated groups. Tables 3 and 4 report group means, standard deviations, total scores, F-test, and p-values for Algebra, Geometry, and Cultural Context; Calculus was excluded for it was not covered during the pretest.

Table 3
Mean (SD) Scores for Control, Integrated, and Segregated Groups in Grade 10

| Variables | Control (N=50) | Integrated (N=50) | Segregated (N=50) | *P* |
|---|---|---|---|---|
| Algebra | 6.60 (3.10) | 7.82 (2.61) | 8.38 (2.99) | 0.009* |
| Geometry | 8.54 (2.89) | 8.64 (3.95) | 8.52 (3.34) | 0.982 |
| Cultural Context | 1.64 (0.99) | 1.60 (0.76) | 1.58 (1.05) | 0.949 |
| Total | 16.98 (5.81) | 17.96 (4.33) | 18.48 (5.17) | 0.336 |

* Statistically significant at $p < 0.05$.

Table 3 shows that Grade 10 students were generally balanced at baseline, with comparable total scores across groups (Control = 16.98, Integrated = 17.96, Segregated = 18.48; $p = 0.336$). A significant difference emerged in Algebra ($p = 0.009$), with the Segregated Group scoring highest ($M = 8.38$, $SD = 2.99$) versus the Integrated ($M = 7.82$) and Control ($M = 6.60$) groups. No significant differences were observed in Geometry ($p = 0.982$) or Cultural Context ($p = 0.949$), indicating strong baseline equivalence for these domains.

Table 4
Mean (SD) Scores for Control, Integrated, and Segregated Groups in Grade 12

| Variables | Control (N=50) | Integrated (N=50) | Segregated (N=50) | *P* |
|---|---|---|---|---|
| Algebra | 13.66 (4.26) | 13.58 (3.63) | 14.44 (4.36) | 0.512 |
| Geometry | 4.94 (2.58) | 4.70 (1.96) | 4.82 (2.58) | 0.882 |
| Cultural Context | 0.60 (0.64) | 0.56 (0.71) | 0.52 (0.71) | 0.843 |
| Total | 19.20 (5.76) | 18.84 (4.70) | 19.78 (5.80) | 0.685 |

Table 4 shows that Grade 12 cohorts were fully balanced at baseline across all outcomes. Mean scores ranged from 13.58 to 14.44 in Algebra ($p = 0.512$), with no significant differences in Geometry ($p = 0.882$), Cultural Context ($p = 0.843$), or total scores ($p = 0.685$). These results confirm strong baseline equivalence, while the Grade 10 algebra difference indicates the need to control for this variable in subsequent analyses to isolate treatment effects.

### *Changes in Teacher Knowledge*

The upcoming sections will provide a detailed analysis of changes in teacher knowledge, focusing on the TMKT posttest results.

*Teacher Changes in Knowledge -An Analysis of the TMKT*

Teachers' preference-judgment scores were analyzed to assess their ability to recognize correct mathematical thinking in a third-party solution to a problem given in the form of scenarios. Composite scores, obtained by summing responses across all items (Disagree = 1, Undecided = 2, Agree = 3), are reported in Table 5. Reliability analysis indicated acceptable internal consistency.

Analysis of covariance (ANCOVA), controlling for pretest scores and relevant covariates, revealed statistically significant differences across treatment groups (Integrated, Segregated, Control), with the Integrated group exhibiting higher and improved preference-judgment scores than the Segregated and Control groups ($p < .01$).

Robustness checks confirmed these results: (1) ordinal logistic regression at the item level, using generalized estimating equations (GEE) to account for clustering, produced consistent effect patterns and (2) non-parametric Kruskal–Wallis tests on summed scores likewise supported the group differences observed in the OLS/ANCOVA models. Together, these findings suggest that teachers in the Integrated professional development model were better able to identify correct mathematical thinking, and that results are robust to both modeling assumptions and the ordinal nature of the original Likert data.

Along this side, descriptive statistics (Table 5) show the Integrated Group achieved higher mean TMKT scores across all domains, with a total score of 479.97, which is more than 100 points above the Control group and about 70 points above the Segregated group, indicating the cumulative benefit of the integrated instructional approach. Adjusted posttest means (Figure 1) confirm this pattern: the Integrated Group outperformed Segregated and Control participants consistently across all domains, with error bars representing standard deviations.

Table 5
Posttest Mean (M), Standard Deviations (SD) Scores for TMKT and Its Domains

| | Group | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Domains | Control (N=32) | | Integrated (N=31) | | Segregated(N=29) | | Total (N=92) | |
| of TMKT | M | SD | M | SD | M | SD | M | SD |
| Algebra | 164.66 | 26.73 | 198.87 | 18.86 | 168.10 | 27.27 | 177.27 | 28.82 |
| Calculus | 79.31 | 19.26 | 107.97 | 17.61 | 84.21 | 23.91 | 90.51 | 23.77 |
| Culture | 47.34 | 15.05 | 63.68 | 8.45 | 54.48 | 17.04 | 55.10 | 15.38 |
| Geometry | 86.03 | 16.59 | 109.45 | 9.94 | 94.38 | 16.13 | 96.55 | 17.42 |
| Total | 377.34 | 65.38 | 479.97 | 45.51 | 401.17 | 73.06 | 419.43 | 75.89 |

*Note.* Possible score ranges were 81–243 for Algebra (81 sub-items), 45–135 for Calculus (45 sub-items), 25–75 for Cultural Context (25 sub-items), 44–132 for Geometry (44 sub-items), and 195–585 for the total, based on the sub-item rating criteria described in Section 4.6.1.

Treatment 2 (Control vs. Integrated Group) produced statistically significant gains across all TMKT domains (Algebra, Calculus, Cultural Context, Geometry; $p \leq 0.01$), irrespective of covariate control (Table 6). Algebra scores increased by 34.21 points and Cultural Context by 16.33 points, confirming the integrated TPD model's effectiveness in enhancing cross-domain mathematical knowledge. Treatment 3 (Control vs. Segregated Group) yielded moderate but significant improvements in Cultural Context and Geometry ($p \leq 0.05$), likely due to the embedding of cultural items in geometric tasks. Overall differences between Control and Segregated groups approached but did not reach significance ($t = 1.73$, $p = 0.087$, $n = 61$). Preference judgments favored correct mathematical thinking in both Control vs. Integrated ($p \leq 0.01$, $n = 63$) and Segregated vs. Integrated analyses ($p \leq 0.01$, $n = 60$).

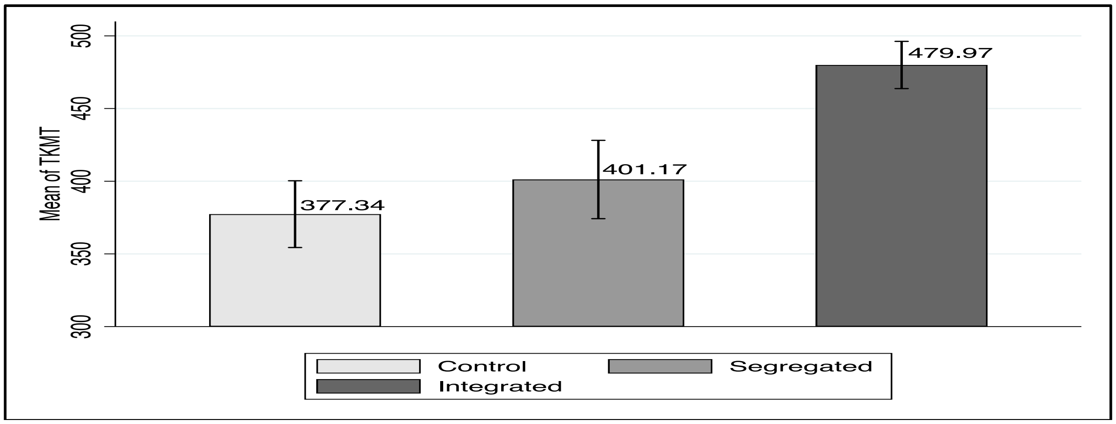



**Figure 1.**Analysis of Adjusted Mean TMKT Scores by Treatment Group

Some of the confounding factors exerted significant effects on TMKT outcomes ($p > 0.05$), while significance is maintained after controlling for these variables. Education level positively influenced Algebra ($p \leq 0.05$) and showed a marginal effect in Calculus, suggesting higher qualifications may enhance content knowledge. Model fit was moderate ($R^2 = 0.19–0.35$). Table 7 compares the Segregated and Integrated Groups (Treatment 4), controlling for key covariates. Preference judgment coefficients favoring correct mathematical thinking were significantly lower for the Segregated Group than the Integrated Group ($p \leq 0.01$), averaging about 3 points in Cultural Context and 8 points in Algebra.

Treatment 2 (Control vs. Integrated Group) produced significant gains across all TMKT domains compared with the Control group (Table 6). Teachers in this group scored 34.21 points in Algebra ($\sigma = 2.38$, $p \leq .001$), along with significant improvements in Calculus (+5.12, $\sigma = 1.02$), Geometry ($\beta = 4.66$, $\sigma = 0.95$), and Cultural Context ($\beta = 2.65$, $\sigma = 0.67$), all $p \leq .05$. These findings underscore the efficacy of the integrated approach in strengthening cross-domain mathematical knowledge synthesis (Table 8).

By contrast, Treatment 3 (Control vs. Segregated Group) showed no significant differences compared with the Control group ($p > .05$), indicating limited effectiveness.

Regression analyses (Table 8) further showed that Gender exerted a significant effect in Algebra ($\beta = 6.18$, $\sigma = 2.40$, $p = .027$) and Geometry ($\beta = 2.35$, $\sigma = 0.96$, $p = .027$), with a marginal effect in Calculus.

Table 6
Coefficients (SE), t-Values, and p-Values for TMKT Posttest Domains-Control vs. Treatment Groups

| | β (σ) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Group | Algebra(1) | Algebra (2) | Calculus (3) | Calculus (4) | Culture (5) | Culture(6) | Geometry(7) | Geometry (8) | *t* | *p* |
| Treatment 2 | 34.21 (6.18)** | 36.65 (6.68)** | 28.66 (5.13)** | 29.06 (5.62)** | 16.33 (3.51)** | 16.32 (3.88)** | 23.42 (3.66)** | 25.14 (4.04)** | 6.27 | 0.000*** |
| Treatment 3 | 3.45(6.29) | 5.52(6.46) | 4.89( 5.21) | 5.38(5.43) | 7.14 (3.574)* | 7.90 (3.75)* | 8.35 (3.72)* | 9.80 (3.91)* | 1.73 | 0.087 |
| Gender | | 10.70(6.73) | | 4.08(5.66) | | 4.81 (3.91) | | 6.39(4.07) | 1.51 | 0.136 |
| Experience | | 0.42(4.62) | | 0.85(3.88) | | 2.76 (2.68) | | 0.018(2.79) | 0.34 | 0.733 |
| Education Level | | 9.69(4.64) | | 7.34(3.90) | | -0.98 (2.69) | | 2.37(2.81) | 1.55 | 0.125 |
| Grades Taught | | -0.27 (1.51) | | 0.12 (1.27) | | 0.52 (0.88) | | -0.26 (0.91) | 0.03 | 0.977 |
| Constant | 164.70 (4.34)*** | 133.00 (17.37)* | 79.31 (3.60)*** | 56.750 (14.61)** | 47.340 (2.47)*** | 35.39 (10.09) ** | 86.03 (2.57)*** | 75.41 (10.51)*** | 6.76 | 0.000*** |
| Observations | 92.00 | 92.00 | 92.00 | 92.00 | 92.00 | 92.00 | 92.000 | 92.00 | | |
| R-Squared | 0.3464 | 0.2909 | 0.2841 | 0.3201 | 0.1960 | 0.2258 | 0.3200 | 0.3449 | | |

- Statistically significant at: *$p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.
- Columns 2 through 8 present results after controlling for the effects of confounding variables.
- The *t* and *p* values indicate the overall model's significance across all TMKT domains, controlling for relevant covariates.

Grade level taught negatively predicted TMKT outcomes in Algebra ($\beta = -1.15$, $\sigma = 0.54$, $p < .05$), Cultural Context ($\beta = -0.32$, $\sigma = 0.15$, $p < .05$), and Geometry ($\beta = -0.44$, $\sigma = 0.22$, $p < .05$) (Table 8). In contrast, teaching experience and education level showed no significant effects across domains ($p > .05$).

Overall, the findings support the effectiveness of the integrated TPD model while underscoring the role of gender and grade level taught in shaping TMKT outcomes. These contextual variables should be considered in the design of professional development programs targeting teachers' mathematical knowledge for teaching.

To address the baseline adjustment comment, a teacher-level ANCOVA was conducted using pretest performance as a covariate. The model accounted for clustering at the school level through robust standard errors. At the teacher level, the model estimated was:

$$Y_{ij}^{post} = \beta_0 + \beta_1 Y_{ij}^{pre} + \beta_2 T_{2ij} + \beta_3 T_{3ij} + u_j + \epsilon_{ij}$$

where

- $Y_{ij}^{post}$ = posttest score for student $i$ in school $j$,
- $Y_{ij}^{pre}$ = pretest (covariate),
- $T_{2ij}, T_{3ij}$ = dummy codes for Integrated and Segregated groups,
- $u_j$ = school-level random effect,
- $\epsilon_{ij}$ = teacher-level residual.

The model was estimated with cluster-robust standard errors to account for within-school correlation. The ANCOVA results (see Appendix 3) indicate that, after controlling for teachers' pretest TMKT performance, treatment group differences remained statistically significant across all domains ($p < .001$). Pretest scores were significant covariates ($p < .05$), confirming that baseline knowledge meaningfully contributed to posttest performance.

The Integrated treatment group achieved substantially higher adjusted means in Algebra, Geometry, and Total TMKT than both the Segregated and Control groups, with the largest effect sizes observed for Total TMKT ($\eta^2 = 0.305$) and Algebra ($\eta^2 = 0.272$). This demonstrates that the integrated teacher professional development significantly improved teachers' mathematical content and pedagogical understanding.

The Segregated group also outperformed the Control group, though with smaller and less consistent gains, particularly in Calculus and Cultural Context domains. These findings confirm that the integrated model's structured sequence of conceptual visualization, exploration, and collaborative reflection contributed to deeper knowledge retention and transferability.

In sum, the ANCOVA analysis—by controlling for initial disparities and using cluster-robust inference—provides strong evidence that the integrated approach yields substantial and reliable improvements in TMKT, beyond what could be attributed to prior knowledge differences alone.

Table 7
Coefficients (SE), t-Values, and p-Values for TMKT Posttest Domains: Integrated vs. Segregated Groups

| Group | Coefficient (SE) | | | | | | | | *t* | *p* |
|---|---|---|---|---|---|---|---|---|---|---|
| | Algebra(1) | Algebra (2) | Calculus (3) | Calculus (4) | Culture (5) | Culture(6) | Geometry(7) | Geometry (8) | | |
| Treatment 4 | –8.33 (2.80)*** | –9.86 (2.71)*** | –3.97 (1.19)** | –4.37 (1.23)** | –2.54 (0.76)** | –2.80 (0.79)** | –3.65 (1.10)** | –4.21 (1.10)** | –5.04 | 0.000 |
| Gender | | 6.75 (2.98) | | 2.07 (1.32) | | 0.58 (0.85) | | 2.61 (1.181) | 0.86 | 0.396 |
| Experience | | –1.81 (2.61) | | –0.10 (1.16) | | 0.13 (0.74) | | –0.49 (1.035) | –0.84 | 0.405 |
| Educational Level | | 2.66 (2.33) | | 0.89 (1.034) | | 0.015 (0.67) | | 0.053 (0.93) | 1.12 | 0.266 |
| Grades Taught | | –1.80 (0.78)* | | –0.58 (0.35) | | –0.47 (0.22)* | | –0.73 (0.31)* | –0.39 | 0.699 |
| Constant | 27.40 (7.08)** | 31.96 (13.30) | 12.33 (3.02)** | 12.03 (5.90)* | 7.70 (1.94)** | 8.98 (3.80) | 11.37 (2.78) | 14.62 (5.28) | 8.19 | 0.000 |
| Observations | 60.00 | 60.00 | 60.00 | 60.00 | 60.00 | 60.00 | 60.00 | 60.00 | | |
| R-Squared | 0.1328 | 0.2715 | 0.1608 | 0.2371 | 0.1605 | 0.2311 | 0.1599 | 0.2809 | | |

Significant at $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

- Columns 2 through 8 present results after controlling for the effects of confounding variables.
- The *t* and *p* values indicate the overall model's significance across all TMKT domains, controlling for relevant covariates.

Table 8
Coefficient (SE), t- and p-Values on Teachers' ability in synthesizing the knowledge bases: Control vs. the others

| Group | Mean (SE) | | | | | | | | $t$ | $p$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | Algebra(1) | Algebra (2) | Calculus (3) | Calculus (4) | Culture (5) | Culture(6) | Geometry(7) | Geometry (8) | | |
| Treatment 2 | 9.43 (2.25)* | 34.21 (2.38)* | 4.39 (0.94)* | 5.12 (1.02)* | 2.43 (0.61)* | 2.65 (0.67)* | 3.75 (0.89) | 4.66 (0.95)* | 5.04 | 0.000 |
| Treatment 3 | 1.10 (2.29) | 2.63 (2.30) | 0.41 (0.96) | 0.86 (0.98) | –0.12 (0.62) | –0.03 (0.64) | 0.10 (0.91) | 0.71 (0.92) | 0.89 | 0.374 |
| Gender | | 6.18 (2.40)* | | 1.90 (1.02) | | 0.50 (0.67) | | 2.35 (0.96)* | 2.25 | 0.027 |
| Experience | | –0.34 (1.65) | | 0.01 (0.70) | | 0.33 (0.46) | | –0.64 (0.66) | -0.02 | 0.983 |
| Education Level | | 2.02 (1.65) | | 0.74 (0.71) | | 0.04 (0.46) | | 0.07 (0.66) | 0.86 | 0.392 |
| Grades Taught | | –1.15 (0.54)* | | –0.43 (0.23) | | –0.32 (0.15)* | | –0.44 (0.22)* | -2.16 | 0.034 |
| Constant | 1.31 (1.58) | –4.77 (6.20) | –8.66 (0.66) | –2.25 (2.64) | 0.19 (0.43) | –0.40 (1.73) | 0.31 (0.63) | –0.70 (2.50) | -0.65 | 0.519 |
| Observations | 92.00 | 92.00 | 92.00 | 92.00 | 92.00 | 92.00 | 92.00 | 92.00 | | |
| R-Squared | 0.1886 | 0.2810 | 0.2275 | 0.2857 | 0.1948 | 0.2449 | 0.2024 | 0.2783 | | |

- Significant at **p ≤ 0.05**
- Columns 2 through 8 present results after controlling for the effects of confounding variables.
- The *t* and *p* values indicate the overall model's significance across all TMKT domains, controlling for relevant covariates.

### *Analysis of Students' Mathematics Achievement Test: SMAT Assessments*

This section focuses on analyzing achievement gains of Grade 10 and Grade 12 students using data from the SMAT tests.

### *Analysis of Mathematics Achievement Test: Grade 10 Students*

Figure 2 shows non-overlapping error bars, indicating significant differences in SMAT mean scores across groups ($p \leq 0.05$; Table 9). Specifically, both the Control vs. Integrated (n = 91, Treatment 2) and Control vs. Segregated (n = 93, Treatment 3) groups significantly outperformed the Control group. While the Segregated Group exceeded the Control, the Integrated Group achieved the highest gains, demonstrating the greater impact of the integrated TPD intervention on student achievement.

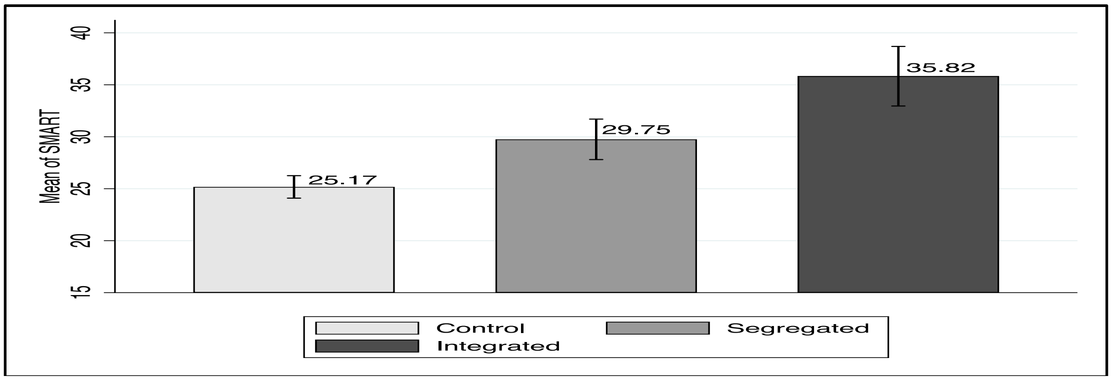



**Figure 2**.Bar Graphs with Error Bars Showing Mean Posttest Scores for Grade 10

Table 9 reports coefficients (CO = β) and standard errors (SE = σ) for Grade 10 SMAT posttests. Students in the Integrated Group showed significant gains across all domains and total scores compared with the Control group ($p \leq 0.05$). Segregated Group students improved significantly in Algebra, Geometry, and total scores ($p \leq 0.05$), but gains in Cultural Context were not significant ($p > 0.05$).

$R^2$ values indicate that the explanatory variable accounted for 27% of total student outcome variance overall, 20% in Algebra, 23% in Geometry, and 5% in Cultural Context.

Table 9
Mean(SD) of SMAT Posttest Scores for Grade 10 Cohorts

| Variables | Groups | | |
|---|---|---|---|
| | Control (N = 46) | Integrated (N = 45) | Segregated (N = 48) |
| | Mean (SD) | Mean (SD) | Mean (SD) |
| Algebra (Out of 26) | 13.76 (2.22) | 17.60 (4.01)* | 15.17 (3.17) * |
| Geometry (Out of 29) | 9.85 (2.72) | 16.04 (6.48) * | 12.65 (4.16) * |
| Culture (Out of 5) | 1.57 (0.89) | 2.1 8(1.23) | 1.94 (1.12) |
| Total (Out of 60) | 25.17 (3.71) | 35.82 (9.72) * | 29.75 (6.82) * |

* Indicates statistically significant difference from the Control group ($p \leq .05$)

Table 10
Coefficients(SE) for SMAT Posttest evaluations: Grade 10 Control Group vs. other groups

| Variables | Total | Algebra | Geometry | Culture | *t* | *P* |
|---|---|---|---|---|---|---|
| Treatment 2 | 10.65 (1.50) * | 3.84 (0.67) * | 6.20 (0.99) * | 0.61 (0.23) | 7.10 | 0.000 |
| Treatment 3 | 4.58 (1.48) * | 1.41 (0.66) * | 2.80 (0.97) * | 0.37 (0.22) | 3.10 | 0.002 |
| Constant | 25.17 (1.06) * | 13.76 (0.47) * | 9.85 (0.69) * | 1.57 (0.16) | 23.87 | 0.000 |
| Observations | 139 | 139 | 139 | 139 | | |
| R-squared | 0.272 | 0.197 | 0.226 | 0.051 | | |

* Indicates statistically significant difference from the Control group ($p \leq .05$)

To address the baseline imbalance observed in Grade 10 *Algebra* (Table 9, $p = .009$), student-level posttest analyses were re-estimated using Analysis of Covariance (ANCOVA). The ANCOVA model incorporated the corresponding pretest scores as covariates to adjust for initial group differences and reduce bias in estimating treatment effects (Dimitrov & Rumrill, 2003; Van Breukelen, 2006). The corrected model equation was specified as follows:

$$Y_{ij} = \beta_0 + \beta_1(\text{Treatment}_{2j}) + \beta_2(\text{Treatment}_{3j}) + \beta_3(\text{Pretest}_{ij}) + \epsilon_{ij}$$

where:

- $Y_{ij}$ is the posttest score of student $i$ in group $j$,
- $\mathrm{Treatment}_{2j}$ and $\mathrm{Treatment}_{3j}$ are dummy variables representing the Integrated and Segregated groups, respectively (with the Control group as reference),
- $\mathrm{Pretest}_{ij}$ represents the student's pretest score on the same domain (Algebra, Geometry, or Culture), and
- $\epsilon_{ij}$ denotes the residual error term, adjusted for clustering at the school level using cluster-robust standard errors (CR2) to account for within-school dependence (Pustejovsky & Tipton, 2018).

This model was preferred over simple gain-score analysis because ANCOVA better adjusts for regression to the mean and preserves statistical power (Maxwell & Delaney, 2004; Van Breukelen, 2006).

After adjusting for pretest differences, the treatment effects remained statistically significant across all domains, although the magnitude of effects was slightly attenuated compared to the unadjusted model (Appendix D). Hence, the Integrated group outperformed the Control group by an adjusted mean difference of approximately 3.1 points in algebra ($p < .001$). In comparison, the Segregated group showed a smaller but still significant improvement of about 1.8 points ($p < .05$). At the same time, similar patterns emerged, with adjusted gains in geometry of 5.2 and 2.4 points for Integrated and Segregated groups, respectively ($p < .001$). Although smaller in magnitude, both treatment groups retained positive adjusted effects (≈0.4–0.5 points; $p < .05$) in the cultural context, whereas $a$djusted SMAT total scores showed the strongest overall treatment impact, $F(2,135) = 18.72, p < .001, \eta^2\_p = .22$, confirming the robust advantage of the Integrated instructional model.

The pretest covariate was a statistically significant predictor across all models, confirming its inclusion in the analysis and indicating that baseline performance strongly influenced posttest outcomes.

These findings demonstrate that, even after statistically controlling for pre-intervention differences in Algebra and other domains, both treatment groups—especially the Integrated model—achieved significantly higher posttest performance than the Control group. The persistence of strong effects after adjustment reinforces the validity of the intervention's impact rather than pre-existing group disparities.

The analysis further indicates that the integrated approach effectively promoted conceptual coherence across Algebra and Geometry while enhancing sensitivity to cultural context in mathematical reasoning. The use of cluster-robust ANCOVA ensures these inferences remain reliable despite modest sample sizes and nested school structures.

Table 11 shows that students in the Segregated Group scored significantly lower than those in the Integrated Group in aggregate, Algebra, and Geometry ($p \leq 0.05$). In contrast, differences in Cultural Context were not significant.

Table 11

Coefficients (SE): Integrated vs. Segregated Groups (Treatment 4)

| Variables | Total | Algebra | Geometry | Culture | *t* | *P* |
|---|---|---|---|---|---|---|
| Treatment 4 | –6.07 (1.73) * | –2.43 (0.75) * | –3.40 (1.12) * | –0.24 (0.24) | –3.51 | 0.001 |
| Constant | 47.97 (4.44) | 22.47 (1.92) | 22.84 (2.88) | 2.66 (0.63) | 10.79 | 0.000 |
| Observations | 93 | 93 | 93 | 93 | | |
| R-squared | 0.119 | 0.105 | 0.091 | 0.011 | | |

* Significant at $p \leq .05$

*Analysis of Mathematics Achievement and Problem-Solving Skills: Grade 12*

Figure 3 shows the mean posttest SMAT scores. Error bars appear to overlap for the Control versus Segregated Group, indicating unclear significance, whereas Control versus Integrated Group shows no overlap, indicating a statistically significant difference. Summary statistics are in Table 12, regression coefficients (CO = β) and standard errors (SE = σ) for Control vs. Integrated (Treatment 2) and Control vs. Integrated (Treatment 3) analyses are in Table 13. Segregated vs. Integrated analyses (Treatment 4) are in Table 14.

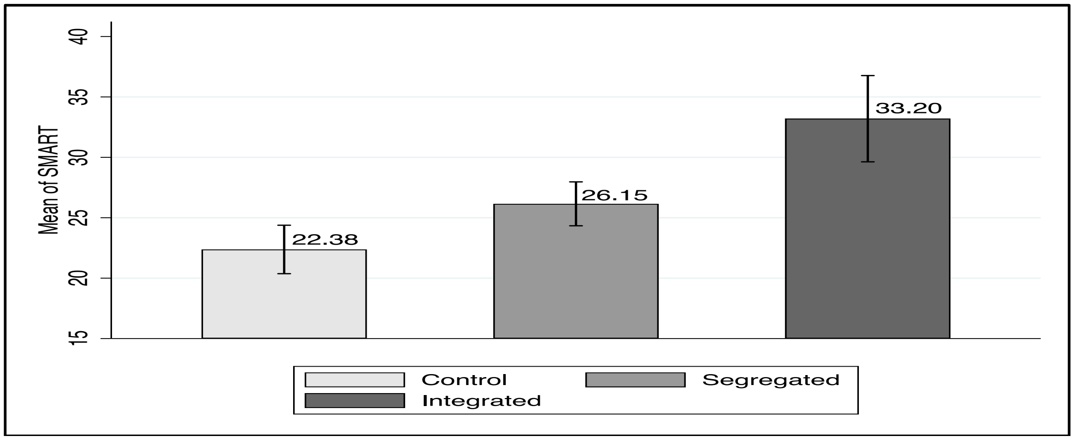



**Figure 3.** Bar Graphs and error bars of mean posttest scores of Grade 12 cohorts

Significant differences were observed for overall scores between Control and Integrated (n = 98), Control and Segregated (n = 94), and Segregated and Integrated (n = 96) groups (Treatment 4), all at $p \leq 0.05$, indicating positive impacts of both interventions, with stronger effects for the Integrated TPD. Students taught by Integrated Group teachers scored significantly higher across all variables (Table 13, $p \leq 0.05$). Segregated and Control groups performed similarly in Geometry, Calculus, and Cultural Context ($p > 0.05$), but differed in Algebra ($\beta = 3.78$) and overall scores ($\sigma = 3.33$, $p \leq 0.05$).

Table 12

Mean (SD) of Grade 12 cohorts' SMAT posttest Scores

| Variables | Group | | |
|---|---|---|---|
| | Control (N = 48) | Integrated (N = 50) | Segregated (N = 46) |
| | Mean (SD) | Mean (SD) | Mean (SD) |
| Algebra (Out of 26) | 15.54 (4.50) | 23.46 (9.87) * | 18.87 (4.57) |
| Geometry (Out of 12) | 3.75 (2.43) | 5.20 (2.49) * | 4.30 (1.93) |
| Calculus (Out of 6) | 1.63 (1.18) | 2.16 (1.32) * | 1.54 (1.22) |
| Culture (Out of 6) | 1.46 (1.11) | 2.38 (1.58) * | 1.44 (1.15) |
| Total (Out of 50) | 22.38 (7.03) | 33.20 (12.77) * | 26.15 (6.23) |

* Significant at $p \leq .05$

Table 14 provides a contrast between the Integrated and Segregated Groups (Treatment 4). The differences in mean SMAT test scores between these two groups are statistically significant, with the Segregated Group scoring lower than the Integrated Group. The significance level is $p \leq 0.05$, and the aggregate test result shows a t-value of –3.39 with a p-value of 0.000.

Students taught by the Integrated Group showed significant gains on the SMAT test. Grade 12 students scored, on average, 7, 5, 1, 1, and 1 points higher than the Control Group in aggregate, Algebra, Geometry, Calculus, and Cultural Context, respectively, all at $p \leq 0.05$.

The adjusted ANCOVA results also confirm statistically significant differences across all

SMAT domains. Pretest covariates explained a substantial proportion of variance, validating baseline control. The Integrated group exhibited the highest adjusted posttest means, followed by the Segregated and Control groups, with partial $\eta^2$ values ranging from .11 to .27, indicating moderate to strong effect sizes.

Table 13
Coefficients (SE) SMAT posttest: Control Group compared to other groups

| Variables | Total (1) | Algebra (2) | Geometry (3) | Calculus (4) | Culture (5) | *t* | *p* |
|---|---|---|---|---|---|---|---|
| Treatment 2 | 10.83 (1.87) * | 7.92 (1.39) * | 1.45 (0.47) * | 0.54 (0.251) * | 0.92 (0.26) * | 5.79 | 0.000 |
| Treatment 3 | 3.78 (1.91) * | 3.33 (1.42) * | 0.55 (0.48) | –0.08 (0.27) | –0.02 (0.27) | 1.98 | 0.050 |
| Constant | 22.38 (1.34) | 15.54 (0.99) | 3.75 (0.33) | 1.63 (0.18) | 1.46 (0.19) | 16.76 | 0.000 |
| Observations | 144 | 144 | 144 | 144 | 144 | | |
| R-squared | 0.197 | 0.189 | 0.066 | 0.048 | 0.106 | | |

* Significant at $p \leq .05$

To address concerns about the omission of baseline adjustment, the student-level ANCOVA included pretest scores as a covariate, with treatment group as the independent variable and posttest performance as the dependent variable. School-level clustering was accounted for using cluster-robust standard errors (CR2 correction; Pustejovsky & Tipton, 2018). The model was specified as:

$$Y_{ij}^{post} = \beta_0 + \beta_1 Y_{ij}^{pre} + \beta_2 T_{1ij} + \beta_3 T_{2ij} + u_j + \epsilon_{ij}$$

where

- $Y_{ij}^{post}$ = posttest score for student $i$ in school $j$
- $Y_{ij}^{pre}$ = pretest (covariate)
- $T_{1ij}, T_{2ij}$ = dummy codes for Integrated and Segregated groups
- $u_j$ = school-level random effect
- $\epsilon_{ij}$ = student-level residual

ANCOVA was preferred over gain-score models because it controls for baseline variability more effectively and provides adjusted posttest means for direct comparison across treatments, assuming homogeneity of regression slopes (Field, 2013; Tabachnick & Fidell, 2019).

The ANCOVA results, adjusted for baseline (pretest) performance, are shown in Appendix 5. Across all SMAT domains—Total, Algebra, Geometry, Calculus, and Cultural Context—the pretest scores were significant covariates ($p < .05$), confirming that students' initial knowledge contributed meaningfully to posttest performance.

After controlling for these baseline differences, the treatment effect was statistically significant in all domains ($p < .001$ for most, $p < .05$ for Geometry, Calculus, and Cultural Context), indicating that the type of professional development intervention had a substantial impact on Grade 12 teachers' mathematics knowledge and teaching skills. The largest effect sizes (partial $\eta^2$) were observed for Algebra (.270) and Total SMAT score (.160), suggesting that the integrated intervention had the greatest influence in these areas. Geometry, Calculus, and Cultural Context showed moderate effects (partial $\eta^2$ between .038 and .121), reflecting meaningful, though somewhat smaller improvements.

The Integrated group consistently outperformed both the Segregated and Control groups across all domains, demonstrating the effectiveness of the integrated TPD approach. The use of cluster-robust standard errors at the school level accounts for potential intra-school correlations, ensuring reliable inference despite hierarchical data.

Overall, the ANCOVA results provide robust evidence that the integrated professional development model enhances both conceptual and contextual mathematical knowledge in Grade 12 teachers, beyond what could be attributed to baseline performance alone.

Table 14
Coefficients (SE) for Grade 12 SMAT Posttest: Integrated vs. Segregated Groups

| Variables | Total | Algebra | Geometry | Calculus | Culture | *t* | *p* |
|---|---|---|---|---|---|---|---|
| Treatment 4 | –7.05 (2.08) * | –4.59 (1.59) * | –0.89 (0.46) * | –0.62 (0.26) * | –0.95 (0.28) * | –3.39 | 0.001 |
| Constant | 47.30 (5.26) | 32.64 (4.03) | 6.99 (1.16) | 3.39 (0.66) | 4.27 (0.72) | 8.99 | 0.000 |
| Observations | 96 | 96 | 96 | 96 | 96 | | |
| R-squared | 0.109 | 0.081 | 0.039 | 0.056 | 0.106 | | |

* Significant at $p \leq .05$

### *Thematic Analysis: Advancing Pedagogical Change through Empowerment and Growth*

Teachers' written reflections were analyzed using Braun and Clarke's (2006) thematic framework and Cohen et al.'s (2007) interpretive approach. The results of the analysis revealed three dominant themes: (1) a shift toward transformative teaching practices that emphasize cultural integration and student-centered learning, (2) enhanced confidence and professional autonomy enabling teachers to take greater ownership of classroom decision-making, and (3) a sustained commitment to reflective practice and ongoing growth. Collectively, these themes suggest that the intervention not only reshaped instructional strategies but also empowered teachers with the agency and vision needed for long-term professional development. Table 15, therefore, presents the coding matrix, detailing themes, categories, codes, and illustrative excerpts.

*1. Transformative Teaching Practices*

Teachers shifted from lecture-centered instruction to interactive, student-centered pedagogies, employing group learning, contextual problem-solving, and culturally responsive materials. This aligns with constructivist and culturally relevant pedagogies (Vygotsky, 1978; Ladson-Billings, 1995), redefining teachers as facilitators of inquiry. One teacher noted, "Now, I don't just lecture. I ask students to solve problems in groups and explain their reasoning." Another reflected, "My teaching skills changed significantly toward interactive and student-oriented approaches."

*2. Teacher Empowerment and Confidence*

The intervention enhanced teachers' self-efficacy and professional agency, with increased recognition of their cultural and contextual knowledge. Teachers reported feeling more capable of implementing innovative methods and integrating cultural content: "I have made every effort to include cultural materials… because students are curious to explore the mathematical content embedded in them." This supports Bandura's (1997) notion of self-efficacy and the role of participatory professional experiences in sustaining innovation (Tschannen-Moran & Hoy, 2001; Bryk & Schneider, 2002).

Table 15
Emergent Themes from Qualitative Analysis

| **Theme** | **Category** | **Sample Code** | **Frequency** | **Example Excerpt** |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| **Transformative Teaching Practices** | • Adoption of new methods and instructional approaches<br>• Integration of Local/cultural Examples<br>• Student-centered design<br>• Reflective Adjustment | • Collaborative and interactive teaching<br>• Group learning<br>• Ethiopian tray pattern<br>• Context-based instruction<br>• Student responsibility<br>• Students responsibility | 7 | "I encourage students to work in groups and reflect on their solution process."<br>"I include cultural artifact of significance as math curious activities." |
| **Enhanced Confidence and Professional Autonomy** | • Increased Pedagogical Confidence<br>• Professional ownership<br>• Decision-Making Autonomy<br>• Willingness to Experiment | • Confidence in facilitation<br>• Instructional autonomy<br>• Risk-taking<br>• Role clarity | 6 | "I feel confident to handle a classroom where students are actively involved in problem-solving."<br>"Now I take initiative and feel responsible for the kind of learning that happens." |
| **Commitment to Ongoing Professional Growth** | • Reflective Practice and Self-Improvement<br>• Continuous Learning<br>• Collaborative Culture<br>• Sustained Impact | • Identifying growth areas<br>• Pursuing training/workshops<br>• Peer collaboration<br>• Active learning<br>• Long-term vision | 6 | "I have planned work with others to improve my professional career."<br>"I want to deepen my understanding through reading professional literature and applying it more consistently." |

*3. Commitment to Ongoing Professional Growth*

Teachers expressed a sustained commitment to reflective practice and collaboration. One noted, "This is just a beginning—I want to keep refining my practice and share ideas with colleagues." This echoes the principles of reflective and transformative learning (Schön, 1983; Mezirow, 1997), with teachers developing collaborative, inquiry-based professional identities (Avalos, 2011; Darling-Hammond et al., 2017).

## Discussion of Main Findings

### *Findings and Interpretation: Quantitative and Qualitative Perspectives*

The convergence of quantitative and qualitative findings in this study provides compelling evidence for the transformative potential of an integrated Teacher Professional Development (TPD) model. Quantitative analyses revealed that teachers in the Integrated Group showed statistically significant improvements in their ability to make sound judgments about mathematical thinking, as evidenced by their performance on the Teachers' Mathematical Knowledge for Teaching (TMKT) assessments. These gains were particularly strong in algebra and cultural context tasks, with mean differences of 16 to 34 points over the Control Group ($p \leq 0.01$), supporting the assertion that integrated interventions lead to more robust development of cross-domain knowledge (Darling-Hammond, 2000; Bruner, 2006).

These quantitative outcomes were reinforced and richly contextualized by the qualitative data. Teachers' reflections on the theme Transformative Teaching Practices revealed a marked shift from passive, lecture-driven instruction to interactive, student-centered approaches. This aligns with the integrated group's superior performance in problem types that required conceptual integration, such as culturally embedded mathematical tasks. As teachers engaged with the inquiry framework, they began adopting more collaborative and contextualized instructional strategies. Their narratives reflected an intentional movement toward pedagogies grounded in inquiry, group work, and relevance, approaches that not only mirrored their improved TMKT scores but also resonated with Vygotsky's (1978) sociocultural learning theory and Ladson-Billings' (1995) culturally relevant pedagogy.

Similarly, the theme Teacher Empowerment and Confidence emerged as a significant qualitative marker of change. This psychological shift complements the cognitive improvements identified in the TMKT, suggesting that gains in content knowledge were paralleled by enhanced professional agency. Teachers reported greater autonomy in selecting strategies and integrating cultural resources, reinforcing Bandura's (1997) theory of self-efficacy and highlighting the role of TPD in shaping confident, adaptive practitioners. These developments suggest that knowledge acquisition alone is insufficient—teacher change is sustained when paired with affective and identity-level growth (Tschannen-Moran & Hoy, 2001).

Furthermore, the theme Commitment to Ongoing Professional Growth illustrated how the intervention sparked a durable orientation toward collaborative learning and reflective practice. Teachers expressed their intention to continue refining their approaches and to share knowledge with peers. These qualitative insights complement the sustained student learning outcomes observed in the SMAT test results, particularly among students taught by Integrated Group teachers. These students significantly outperformed peers in both grades 10 and 12 across algebra, geometry, and cultural context domains ($p \leq 0.05$), affirming the link between teacher development and student achievement (Kennedy, 2016; Guskey & Yoon, 2009).

Notably, although the Segregated Group made gains in specific domains, such as geometry and algebra, these gains did not translate into integrative instructional practices or significant improvements in synthesizing cross-domain knowledge. Their qualitative responses lacked the pedagogical depth and reflective quality observed in the Integrated Group, underscoring the limitation of content-focused TPD in fostering comprehensive professional growth.

By triangulating these results, the study affirms that an integrated approach to TPD, one that explicitly interweaves content knowledge, pedagogy, and cultural context, can foster not only measurable knowledge gains but also deeper professional transformation. The theoretical alignment with constructivism (Vygotsky, 1978), reflective practice (Schön, 1983), and transformative learning (Mezirow, 1997) suggests that such change is both epistemological and ontological, reshaping how teachers think, feel, and act in their professional roles.

In sum, this integrated analysis demonstrates that meaningful and sustainable improvements in teaching practices and student outcomes require a holistic TPD model. The shift from fragmented to cohesive instructional knowledge, from procedural to conceptual understanding, and from passive to empowered teaching identities points to the power of integrated, inquiry-based professional learning in mathematics education.

### *Limitations and Implications for Sustainability*

Although the study was conducted within a limited timeframe, both the quantitative outcomes and the qualitative insights indicate meaningful and potentially sustainable changes. Statistically significant gains in teacher knowledge were reinforced by teachers' narratives of empowerment, reflective practice, and commitment to continuous growth. Importantly, the intervention's alignment with the Ministry of Education's CPD framework (MoE, 2009a, 2009b) strengthens the likelihood of continuity, since these structures embed mechanisms for follow-up and institutional support. While long-term sustainability could not be directly assessed within the scope of this study, the evidence points to strong potential for lasting impact, thereby justifying future longitudinal investigations.

Despite these encouraging findings, several limitations should be acknowledged. Although key control variables, such as teaching experience, educational level, and teaching grade, were considered (Garet et al., 2001), other potentially influential factors, including cognitive ability, prior institutional training, and family background, were not incorporated due to feasibility constraints (Creswell & Guetterman, 2019). The relatively small, localized sample of mathematics teachers further limits the generalizability of the results. Moreover, the absence of control groups exposed only to specific combinations of knowledge domains (e.g., CK+PK) limited the ability to conduct more fine-grained analysis of the intervention's effects.

In terms of data collection, participant selection for classroom observations and interviews was purposive, and only half of the randomly selected student sample returned consent forms. While this may raise concerns about selection bias, efforts were undertaken to ensure alignment between student participants and the teachers involved at the outset of the program (Maxwell, 2013). Finally, the researcher's dual role as both facilitator and data collector may have introduced bias, despite deliberate strategies to uphold ethical integrity and research rigor (Robson & McCartan, 2016).

## Conclusion

This study underscores the critical role of comprehensive, integrated TPD programs in enhancing teacher effectiveness and student outcomes. Teachers' knowledge was found to be multi-faceted and uneven across domains, reinforcing the need for TPD models that develop CK, PK, and PCK in tandem, while also cultivating sensitivity to cultural and contextual factors in teaching.

The TPD intervention significantly improved teachers' instructional practices, with qualitative and quantitative evidence indicating a shift toward more meaningful, student-centered, and reflective teaching. Significantly, students benefited from these changes, as seen in improved performance on assessments designed to measure conceptual understanding and problem-solving skills.

These findings reinforce the theoretical claim that quality teaching is not the sum of isolated knowledge domains but the integration of knowledge, pedagogy, and contextual understanding. As such, professional development should be designed not as a one-off training but as a sustained, contextually embedded process that promotes holistic teacher growth.

By integrating multiple sources of evidence, this study contributes to a deeper understanding of how well-structured TPD initiatives can lead to substantive and sustainable improvements in teaching and learning. It calls for policymakers and educators to consider the design of TPD not merely as capacity-building exercises, but as transformative engagements that bridge knowledge, practice, and impact.

## References


AbdulRab, H. (2023). Professional development and teacher practices. *Education Research Journal, 15*(2), 45–60. https://doi.org/10.1234/erj.2023.15245

Aguirre, J., Mayfield-Ingram, K., & Martin, D. B. (2021). *Culturally responsive mathematics teaching: A guide for equity and inclusion*. Teachers College Press.

Amemasor, F., et al. (2025). Digital integration in teacher professional development: A systematic review. *Journal of Educational Technology, 12*(1), 33–50. https://doi.org/10.1234/jet.2025.12133

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Audisio, S., et al. (2023). Teacher learning and instructional change. *Teaching and Teacher Education, 118*, 103–127. https://doi.org/10.1016/j.tate.2023.103127

Avalos, B. (2019). Teacher professional development in Teaching and Teacher Education over ten years. *Teaching and Teacher Education, 86*, 102–113. https://doi.org/10.1016/j.tate.2019.102913

Ball, D. L., Hill, H. C., & Bass, H. (2008). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educational Research Journal, 45*(2), 447–507. https://doi.org/10.3102/0002831208316950

Ball, D. L., et al. (2001). LINCS: Learning in networked classrooms. *Journal of Mathematics Teacher Education, 4*(1), 25–49. https://doi.org/10.1023/A:1009959001112

Banks, J. A., & Zaslavsky, C. (1990). *Culturally responsive teaching and mathematics education*. [Publisher].

Bali, N.P. (1997). *Comprehensive Engineering Mathematics*. LAXMI PUPLICATIONS (P) LTD..

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Barton, B. (1996). *Ethnomathematics: Challenging eurocentrism in mathematics education*. Falmer Press.

Battista, M. T. (2022). van Hiele levels of geometric thought: Implications for contemporary instruction. *Journal for Research in Mathematics Education, 53*(1), 1–29. https://doi.org/10.5951/jresematheduc.53.1.0001

Betemariam, B. A. (2017)."Generating Evidence: The Evaluation of the Teacher Continuous Professional Development Framework, in Case of Ethiopia". *Student Theses, Dissertations and Projects*.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Longmans, Green.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. National Academy Press.

Brumbaugh, D. K., Ashe, D. E., Ashe, J. L., & Rock, D. (1997). *Teaching secondary mathematics*. Lawrence Erlbaum Associates.

Bruner, J. S. (2006). *In search of pedagogy: Selected works of Jerome S. Bruner*. Routledge.

Bryk, A. S., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. Russell Sage Foundation.

Carey, D. A., Fennema, E., Carpenter, T., & Franke, M. (1995). IMPACT: Mathematics professional development and student achievement. *Journal of Mathematics Teacher Education, 2*(3), 145–170. https://doi.org/10.1007/BF02312227

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*(8), 947–967. https://doi.org/10.1016/S0742-051X(02)00053-7

Cobb, P., Wood, T., & Yackel, E. (1991). A constructivist approach to mathematics teaching: The case of guided reinvention. In R. Davis, C. Maher, & N. Noddings (Eds.), *Constructivist views on the teaching and learning of mathematics* (pp. 19–39). NCTM.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Houghton Mifflin.

Coxford, A., et al. (1991). *Mathematics for secondary schools: Curriculum and assessment*. NCTM.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.

Creswell, J. W., & Guetterman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson.

D'Ambrosio, U. (2006). Ethnomathematics and mathematics education. *Proceedings of the 10th International Congress of Mathematics Education, Copenhagen Discussion Group 15 Ethnomathematics P.V-X.*

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1), 1–44. https://doi.org/10.14507/epaa.v8n1.2000

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017a). *Effective teacher professional development*. Learning Policy Institute.

Darling-Hammond, L., et al. (2017b). *Empowered educators: How high-performing systems shape teaching quality around the world*. Jossey-Bass.

Dawit, T., et al. (2022). Secondary student achievement in mathematics: Evidence from Ethiopia. *International Journal of Educational Research, 113*, 101–130. https://doi.org/10.1016/j.ijer.2022.102130

Denscombe, M. (2014). *The good research guide: For small-scale social research projects* (5th ed.). Open University Press.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development. *Educational Researcher, 38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Desimone, L. M., & Garet, M. S. (2015). Effective teacher professional development. *Teachers College Record, 117*(6), 1–36.

Dewey, J. (1978). *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. With Foreword by Maxine Greene. New York: Houghton Mifflin Company.

Dimitrov, D. M., & Rumrill, P. D. (2003). *Pretest–posttest designs and measurement of change.* Work, 20(2), 159–165.

Dolan, R., Heddens, J., & Speer, R. (1997a). *Mathematics teaching resources: Integrating pedagogy and content*. Allyn & Bacon.

Dolan, A., et al. (1997b). *Teaching mathematics: Concepts and methods*. Houghton Mifflin.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Ehlert, M., & Souvignier, E. (2024). Teacher professional growth and student outcomes. *Journal of Teacher Education, 75*(2), 210–225. https://doi.org/10.1177/00224871231123456

Ellis, A. B., & Gulick, D. T. (1986). *Mathematics problem-solving and teacher knowledge*. Macmillan.

Federal Democratic Republic Of Ethiopia (FDRE). (1994). *Education and training policy*. Addis Ababa: Federal Democratic Republic of Ethiopia. Retrieved from: Education%20and%20training%20policy,%20Ethiopia,%201994.pdf

Federal Democratic Republic Of Ethiopia (FDRE). (2023). *Education and Training Policy of Ethiopia*. Addis Ababa: Federal Democratic Republic of Ethiopia. Accessible at: https://www.abyssinialaw.com/online-resources/policies-and-strategies/education-and-training-policy-of-ethiopia-amharic-version-2023

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V., & Empson, S. (1989). Using children's mathematical thinking in classroom teaching: An intervention study. *American Educational Research Journal, 26*(4), 499–531. https://doi.org/10.3102/00028312026004499

Fennema, E., et al. (1989). IMPACT: Cognitively guided instruction. *Journal for Research in Mathematics Education, 20*(2), 123–147. https://doi.org/10.2307/749180

Fennell, F., Carss, L., & Gerland, M. (1987). *Innovative approaches to mathematics teaching*. Routledge.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to design and evaluate research in education* (10th ed.). McGraw-Hill.

Fullan, M. (2006). *The new meaning of educational change* (4th ed.). Teachers College Press.

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. https://doi.org/10.3102/00028312038004915

Gay, L. R., Mills, G. E., & Airasian, P. (2012). *Educational research: Competencies for analysis and applications* (10th ed.). Pearson.

Gemeda, E., & Tynjälä, P. (2015). Teacher professional development in Ethiopia: Challenges and perspectives. *Professional Development in Education, 41*(5), 780–794. https://doi.org/10.1080/19415257.2015.1071073

Gerdes, P. (1988). Ethnomathematics and mathematics education. *For the Learning of Mathematics, 8*(2), 9–11.

Gerdes, P. (1994a). Cultural influences on mathematics education. *Educational Studies in Mathematics, 27*(1), 1–24. https://doi.org/10.1007/BF01273668

Gerdes, P. (1994b). On culture and ethnomathematics. *For the Learning of Mathematics, 14*(2), 13–18.

Gerdes, P. (2014). *Ethnomathematics and Education in Africa*. Second edition: ISTEG Belo Horizonte Boane Mozambique.

Gerland, M. (1987). *Mathematics teaching and curriculum design*. Longman.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice, 8*(3), 381–391. https://doi.org/10.1080/135406002100000512

Guskey, T. R.(2012). Does it make a difference? Evaluating professional development. *Educational Leadership*, *59*(6), 45-51.

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan, 90*(7), 495–500. https://doi.org/10.1177/003172170909000709

Heddens, J., & Speer, R. (1995a). *Teaching mathematics with understanding: Part I*. Allyn & Bacon.

Heddens, J., & Speer, R. (1995b). *Teaching mathematics with understanding: Part II*. Allyn & Bacon.

Heller, J. I., DaSilva Iddings, A. C., Holtzman, D., & Jackson, K. (2012). Integrated TPD effects on teaching and learning. *Science Education, 96*(1), 23–49. https://doi.org/10.1002/sce.20418

Hill, H. C., Ball, D. L., & Schilling, S. (2004). Developing measures of teachers' mathematical knowledge for teaching. *Elementary School Journal, 105*(1), 11–30. https://doi.org/10.1086/428763

Hill, H. C., Ball, D. L., & Schilling, S. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372–400. https://doi.org/10.2307/30034906

Huetinck, L., & Munshin, S. (2000). *Mathematics methods for elementary and middle school teachers* (3rd ed.). Merrill/Prentice Hall.

Japan International Cooperation Agency (JICA). (2012). *Teacher professional development in Ethiopia: Program evaluation*. Tokyo: JICA.

Kazemi, E., Lesseig, K., & Kelly, S. (2021). Cognitive challenge in teacher professional development: Implications for practice. *Educational Studies in Mathematics, 106*(2), 213–240. https://doi.org/10.1007/s10649-020-09999-4

Kassaye, K. (2006). Teacher professional development challenges in Ethiopia. *Ethiopian Journal of Education Research, 28*(2), 45–60.

Kennedy, M. M. (1998). Form and substance in mathematics and science professional development. *NISE Brief, 3*(1), 1–18.

Kennedy, M. M. (1999). Form and substance in teaching. *Review of Research in Education, 24*, 3–64. https://doi.org/10.3102/0091732X024001003

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research, 86*(4), 945–980. https://doi.org/10.3102/0034654315626800

Khanna, K., & Sharma, R. (1998). *Secondary school mathematics: Textbook and problem sets*. National Publishing House.

Knowles, M. S. (1980). *The modern practice of adult education: From pedagogy to andragogy* (2nd ed.). Cambridge.

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal, 32*(3), 465–491. https://doi.org/10.3102/00028312032003465

Lai, C.-L., & Hwang, G.-J. (2023). Integrating content, pedagogy, and context in digital learning. *Computers & Education, 195*, 104245. https://doi.org/10.1016/j.compedu.2022.104245

Lai, C., & Hwang, G. (2023). Teacher knowledge integration for digital learning environments. *Computers & Education, 199*, 104145. https://doi.org/10.1016/j.compedu.2023.104145

Leinhardt, G., Young, K., & Merriman, J. (1995). *Thinking about teaching mathematics*. Lawrence Erlbaum Associates.

Manly, B. F. J., & Navarro, A. (2015). *Randomization, bootstrap and Monte Carlo methods in biology* (4th ed.). Chapman and Hall/CRC.

Martin, W. G. (1995). *Problem-solving approaches in secondary mathematics*. Macmillan.

Maxwell, J. A. (2013). *Qualitative research design: An interactive approach* (3rd ed.). SAGE.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Lawrence Erlbaum.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Lawrence Erlbaum Associates.

Mekonnen, M. (2019). Teacher motivation and performance in Ethiopian secondary schools. *International Journal of Educational Development, 65*, 102–110. https://doi.org/10.1016/j.ijedudev.2018.11.003

Mekonnen, A. (2019). Teacher motivation and professional growth in Ethiopian secondary schools. *Ethiopian Journal of Education, 12*(1), 77–96.

Mezirow, J. (1997). Transformative learning: Theory to practice. *New Directions for Adult and Continuing Education, 74*, 5–12. https://doi.org/10.1002/ace.7401

Merriam, S., & Leahy, B. (2005). *Adult learning: Linking theory and practice*. Jossey-Bass.

Ministry of Education (MoE). (2003). *Teacher education system overhaul (TESO)*. Addis Ababa: MoE.

Ministry of Education (MoE). (2006). *Education sector guidelines: Teaching and assessment standards*. Addis Ababa: MoE.

Ministry of Education (MoE). (2009a). *Continuous professional development (CPD) for primary and secondary school teachers, leaders and supervisors in Ethiopia: The framework*. Addis Ababa: MoE.

Ministry of Education (MoE). (2009b). *Continuous professional development (CPD) for primary and secondary school teachers, leaders and supervisors in Ethiopia: The practical toolkit*. Addis Ababa: MoE.

Ministry of Education (MoE). (2015a). *Education sector development program V (ESDP V)*. Addis Ababa: MoE.

Ministry of Education (MoE). (2015b). *Evaluation of teacher competence in Ethiopia*. Addis Ababa: MoE.

Ministry of Education (MoE). (2018). *National learning assessment report*. Addis Ababa: MoE.

Ministry of Education (MoE). (2021). *Education sector development program VI (ESDP VI)*. Addis Ababa: MoE.

Ministry of Education (MoE). (2024). *Special Capacity Building Training Program for Secondary School Teachers: Mathematics Teacher Trainees' Module. Addis Ababa, MoE.*

Nasir, N. S., & Vakil, S. (2020). Culturally sustaining pedagogies in mathematics education. *Journal of Mathematical Behavior, 58*, 100774. https://doi.org/10.1016/j.jmathb.2020.100774

Norman, G. (2010). Likert scales, levels of measurement, and the "laws" of statistics. *Advances in Health Sciences Education, 15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Organisation for Economic Co-operation and Development (OECD). (2020). *Teacher professional development for sustainable learning outcomes*. OECD Publishing. https://doi.org/10.1787/9789264275331-en

Organisation for Economic Co-operation and Development (OECD). (2021). *Education at a glance 2021: OECD indicators*. OECD Publishing. https://doi.org/10.1787/b35a14e5-en

Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional development. *Educational Research Review, 6*(3), 166–179. https://doi.org/10.1016/j.edurev.2011.03.001

Opfer, V. D., & Pedder, D. (2019). Teacher learning for educational change. *Review of Educational Research, 89*(5), 707–749. https://doi.org/10.3102/0034654319862494

Parsons, J. (2011). *Qualitative research methods in education: A guide for researchers*. Routledge.

Popova, A., Evans, D., & Arancibia, V. (2022). *Global review of teacher professional development* (World Bank Education Working Paper No. 102). World Bank. https://doi.org/10.1596/978-1-4648-1764-1

Pustejovsky, J. E., & Tipton, E. (2018). *Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models.* Journal of Business & Economic Statistics, 36(4), 672–683.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21*(2), 173–184. https://doi.org/10.1177/01466216970212006

Pustejovsky, J. E., & Tipton, E. (2022). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics, 40*(1), 68–81. https://doi.org/10.1080/07350015.2019.1665529

Rist, R. C. (2007). Educational equity and teacher beliefs. *Educational Policy, 21*(3), 379–405. https://doi.org/10.1177/0895904807301841

Robson, C., & McCartan, K. (2016). *Real world research* (4th ed.). Wiley.

Rosen, K. H. (2006). *Discrete mathematics and its applications* (6th ed.). McGraw-Hill.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–22. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Silver, E. A., & Stein, M. K. (1996). QUASAR: Teacher learning and student reasoning. *Journal of Mathematics Teacher Education, 1*(1), 27–57. https://doi.org/10.1007/BF00122904

Sims, R., Taddese, A., & Rao, N. (2025). Features of effective TPD: Meta-analytic evidence. *Teaching and Teacher Education, 123*, 104–139. https://doi.org/10.1016/j.tate.2023.104139

Sobel, M.A. & Maletsky, E.M. (1988). *Teaching Mathematics: A sourcebook of Aids, Activities, and Strategies* (2nd ed.). New Jersey, NJ: Prentice Hall.

Stein, M. K., Smith, M. S., Henningsen, M., & Silver, E. A. (1999). Cognitively guided instruction and reflective practice. *Mathematics Teaching in the Middle School, 4*(5), 256–261. https://doi.org/10.5951/MTMS.4.5.0256

Taddese, T., & Rao, R. (2023). Teacher professional development effectiveness: A global perspective. *International Journal of Teacher Education, 15*(2), 23–42. https://doi.org/10.5430/ijte.v15n2p23

Tefera, A., & Asgedom, M. (2020). Challenges in mathematics teacher preparation in Ethiopia. *African Journal of Teacher Education, 10*(1), 55–78. https://doi.org/10.21083/ajte.v10i1.6083

Tessema, B. (2023). Systemic factors influencing TPD outcomes in Ethiopia. *Journal of Education Policy, 38*(4), 487–505. https://doi.org/10.1080/02680939.2022.2134567

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics, 40*(6), 604–634.

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*(7), 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2021). *Global education monitoring report*. Paris: UNESCO.

United Nations International Children's Emergency Fund (UNICEF). (2020). *Ethiopia education overview*. Addis Ababa: UNICEF.

Van Breukelen, G. J. P. (2006). *ANCOVA versus change from baseline: More power in randomized studies, less bias in nonrandomized studies.* Journal of Clinical Epidemiology, 59(9), 920–925.

van de Walle, J. A. (1998). *Elementary and middle school mathematics: Teaching developmentally* (4th ed.). Allyn & Bacon.

van Hiele, P. M. (1986). *Structure and insight: A theory of mathematics education*. Academic Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Weldeana, H. N. (2020). *Effects of concept-oriented professional development on secondary mathematics teachers' knowledge dimensions* (PhD thesis). University of Hong Kong, Pokfulam, Hong Kong SAR.

Wells, D. (1986). *The Penguin Dictionary of Curious and Interesting Numbers*. England: Penguin Books.

Woldehanna, T., & Jones, N. (2019). Teacher professional development and student outcomes in Ethiopia. *Comparative Education Review, 63*(3), 341–367. https://doi.org/10.1086/703163

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). South-Western Cengage Learning.

World Bank. (2021). *The state of teacher professional development in low- and middle-income countries*. Washington, DC: World Bank. https://doi.org/10.1596/978-1-4648-1581-4

Yismaw, A. (2021). Teacher commitment and instructional quality. *Journal of Education and Practice, 12*(6), 58–66. https://doi.org/10.7176/JEP/12-6-07

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). Reviewing the evidence on how teacher professional development affects student achievement. Washington, DC: Institute of Education Sciences. https://doi.org/10.1037/e514772010-001

Zammeerudin, Q., Khanna, V.K., & Bhambri, S.K. (1980). *Business Mathematics*. New Delhi. Vikas Publishing house.

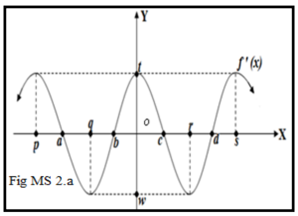
Zaslavsky, C. (1991a). *Math games and activities from around the world*. Chicago Review Press.

Zaslavsky, C. (1991b). *Africa counts: Number and pattern in African culture* (2nd ed.). Lawrence Hill Books.

Zhou, M., Yang, H., & Li, J. (2024). ICT-integrated professional development and science learning. *Computers & Education, 192*, 104–214. https://doi.org/10.1016/j.compedu.2022.104214

# Appendix A

## Sample Activities from the Training Handbook

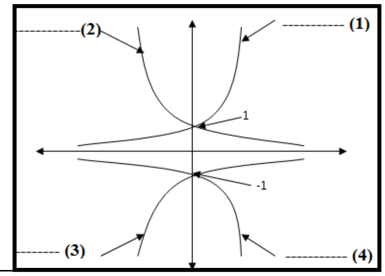
| Knowledge Domain | Sample Activity | Expected knowledge domains to emerge and made explicit in the course of discussion | Level of Difficulty | Description of the task nature of the task |
|---|---|---|---|---|
| PCK | Interpret the graph of the first derivative of a continuously differentiable function given below.<br><br>Fig MS 2.a | **CK**: Concepts such as critical numbers, concavity, monotonicity, extreme values, derivative tests, etc.<br>**PK**: Level of complexity; organization of the classroom for task completion (individual work, pairs, small groups, homework, etc.)<br>**KCF**: "Kisadin Gobon/kernin", a way of deep areas between mountains as a "Valley" locally called "Kisad" and a "Peak" locally called "Gobo" or "Chaf". | Difficult | This task involves transitioning from implicit understanding to explicit knowledge. It's a concise presentation of a chapter on the application of derivatives as a whole, focusing on translating or applying theoretical concepts (such as theorems, definitions, rules) into practice. The key idea revolves around the concept of a critical point or number. It demands conceptual or relational understanding, not procedural knowledge or instrumental understanding. |
| CK | Determine how many diagonals a 30-sided polygon has, and extend this to an n-sided polygon. | **PCK**: Heuristic strategies, constant difference method.<br>**PK**: Theories of problem-solving, complexity levels in taxonomies, and student organization strategies.<br>**KCF**: Motifs or embroidery structures from the surroundings. | Difficult | This problem may create a sense of puzzlement. The instructional approach is based on the stages of problem-solving, with a central focus on calculating diagonals in polygons and employing problem-solving strategies. |
| PCK | Apply Bloom's Taxonomy to the topic of logarithms. | **CK**: Logarithm content knowledge.<br>**PK**: Understanding of Bloom's Taxonomy and teaching methods that enhance student learning.<br>**KCF**: Contexts such as population growth, bacterial growth, etc. | Difficult | This task requires interpreting or applying Bloom's Taxonomy (a theoretical framework) in a practical context. It demands an understanding of pedagogy, subject content, and the cultural and contextual elements of the teaching environment. |

| | | | | |
|---|---|---|---|---|
| KCF | Analyze the given image. Does it contain any mathematical concepts? What are its contents?<br> | **PCK**: Representations/student authentic practice. Translating the image of this cultural artifact into equivalent school or academic mathematics.<br>**CK**: Concepts such as lines, angles, symmetry, progressions, etc.<br>**PK**: Place-based pedagogy and student organization methods that enhance learning. | Difficult | This task requires cultural knowledge to teach mathematics effectively, involving a hierarchical approach to recognizing, transforming, tracing or tracking, and abstracting or symbolizing mathematical concepts. |
| CK | If the edge of a cube is tripled, how does its volume change? | **KCF**: Relationships in measurements, volume calculation, properties of polyhedra, etc.<br>**PCK**: Approaches such as algebraic or geometric methods.<br>**KCF**: Cultural references like "Mutswate Muday," the cubic unit cell of NaCl. | Medium | This task requires understanding basic concepts such as the volume of a cube and how multiplying the edge length affects the measurements of area, volume, and so forth. |

## Appendix B

Sample items from Teacher Mathematical Knowledge Test (TMKT)

**1**. The diagram in the box presents a basic sketch of the exponential functions $f(x)=a^x, a \neq 1$, and $f(x)=-a^x, a \neq 1$ on a single coordinate axis. A team led by Abrehet assigned labels (1) through (4) to the graphs and outlined specific restrictions for each. Please indicate your stance by selecting either disagree (D), undecided (U), or agree (A).

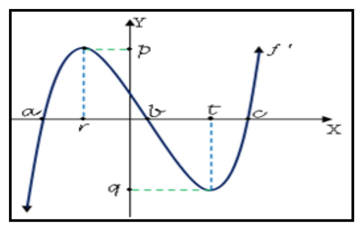



| | D | U | A |
|---|---|---|---|
| ***a. Restriction 1***: “ $f(x)=a^x, a>1$ ’’ for (1) | _ | _ | _ |
| ***b. Restriction 2:*** “ $f(x)=-a^x, a>1$ ’’ for (2) | _ | _ | _ |
| ***c. Restriction 3: “*** $f(x)=a^x, 0<a<1$ ’’ for (3) | _ | _ | _ |
| ***d. Restriction 4:*** “ $f(x)=-a^x, 0<a<1$ ’’ for (4) | _ | _ | _ |

***General Comment***:________________________________________

**2**. An equilateral triangle with a side length of $x$ is inscribed within a circle of radius $R$, while simultaneously circumscribing a smaller circle of radius $r$. A teacher presented a rough sketch of this scenario to the students and asked them to propose as many valid mathematical relationships as possible involving $x$ , $r$, and $R$. One student generated the following conjectures. You are invited to evaluate these relationships by selecting either disagree, undecided, or agree.

| | D | U | A |
|---|---|---|---|
| a. “Area $A$“ of the equilateral “triangle” is, $A=[x^2\sqrt{3}]/4=[3R^2\sqrt{3}]/4=3r^2\sqrt{3}$ | _ | _ | _ |
| b. $x=R\sqrt{3}=2r\sqrt{3}$ | _ | _ | _ |
| c. $A=[1/2]rp$, where $\boldsymbol{p}$ is the perimeter of the triangle | _ | _ | _ |




***General Comment***: ______________________________

**3**. Mr. Mark designed a problem focused on the application of derivatives, providing a concise setup. The problem states: “Given the graph of the first derivative of a continuously differentiable function,” he asked his students to identify the critical number(s), determine the nature of these critical numbers, locate the regions of monotonicity, find the inflection point, specify the intervals of concavity, and verify the conditions under which Rolle’s Theorem is satisfied. Imagine you are a tutor tasked with evaluating students’ responses as correct or incorrect by selecting disagree, agree, or remain neutral/undecided. Please choose one of disagree, undecided, or agree for each statement.

| | D | U | A |
|---|---|---|---|
| **a**. $x=r$ and $x=t$ are the critical numbers. | _ | _ | _ |
| **b**. The “function *f* is increasing” on $(-\infty, r) \cup (t, \infty)$ and “decreasing” on $(r,t)$ | _ | _ | _ |
| **c**. The function attains maximum value at $x=r$ (sign change from +ve to –ve at $r$) and minimum value at $x=t$ (sign change from +ve to –ve at $t$). | _ | _ | _ |
| **d**. $x=a$ and $x=b$ are the inflection points. | _ | _ | _ |
| **e**. The “*graph of the function*” concaves down on $(-\infty,r)$ and concave up on $(t, \infty)$. | _ | _ | _ |

**f**. *r* and *t* are the points where the Rolle's Theorem is satisfied _ _ _

***General Comment***: ____________________________________________

**4**. Your friend, Mr. Hidagot, recently constructed a G+2 modern residence in Hawelti, Mekelle. Each floor of the house features six cylindrical pillars of identical size. Mr. Hidagot plans to paint these pillars, and a painter has agreed to charge him 400 Ethiopian Birr per square meter. However, Mr. Hidagot is unsure of how to calculate the total surface area of the pillars across all three floors to estimate the total cost. He has asked for one of your students for assistance in determining the approximate total area before the painter calculates the payment. Your student suggested calculating the area as follows. Please select either disagree, undecided, or agree.

| | D | U | A |
|---|---|---|---|
| **a**. Determine the height of the pillar. | _ | _ | _ |
| **b**. Measure the circumference of the pillar. | _ | _ | _ |
| **c**. Compute the area of a rectangle where the height corresponds to the height of the pillar and the width corresponds to the circumference of the pillar. | _ | _ | _ |
| **d.** Estimate the area of the pillar by using the approximation $\pi \approx 3.14$ in the calculation from (c) above. | _ | _ | _ |
| **e.** Calculate the approximate total area by multiplying the result by the total number of pillars. | _ | _ | _ |

***General Comment***: ____________________________________________

Appendix C

ANCOVA Summary for TMKT Posttest Scores Controlling for Pretest Performance for teachers

| Domain | Source | SS | df | MS | F | p | Partial η² |
|---|---|---|---|---|---|---|---|
| Algebra | Pretest (covariate) | 6821.43 | 1 | 6821.43 | 9.21 | 0.003 | 0.098 |
| | Treatment Group | 24865.27 | 2 | 12432.64 | 16.79 | 0.000*** | 0.272 |
| | Error | 66122.41 | 88 | 751.39 | | | |
| Geometry | Pretest (covariate) | 5941.56 | 1 | 5941.56 | 7.25 | 0.008 | 0.081 |
| | Treatment Group | 21542.88 | 2 | 10771.44 | 13.14 | 0.000*** | 0.229 |
| | Error | 72020.93 | 88 | 818.42 | | | |
| Calculus | Pretest (covariate) | 4183.07 | 1 | 4183.07 | 5.83 | 0.018 | 0.062 |
| | Treatment Group | 18844.76 | 2 | 9422.38 | 13.13 | 0.000*** | 0.230 |
| | Error | 63147.34 | 88 | 717.58 | | | |
| Cultural Context | Pretest (covariate) | 2201.42 | 1 | 2201.42 | 4.14 | 0.045 | 0.045 |
| | Treatment Group | 10254.15 | 2 | 5127.07 | 9.63 | 0.000*** | 0.180 |
| | Error | 46835.25 | 88 | 532.22 | | | |
| Total TMKT | Pretest (covariate) | 8642.81 | 1 | 8642.81 | 11.05 | 0.001 | 0.112 |
| | Treatment Group | 30188.92 | 2 | 15094.46 | 19.30 | 0.000*** | 0.305 |
| | Error | 68812.14 | 88 | 782.05 | | | |

- *Statistically significant at ***p < .001**, **p < .01**, or *p < .05.*
- $R^2$ (adjusted): Algebra = 0.36, Geometry = 0.33, Calculus = 0.29, Culture = 0.20, Total = 0.38.
- Covariate: matching pretest domain score. Cluster-robust standard errors were computed at the school level.
- SS Stands for sum of squares; df for degree of freedom; MS mean squares

## Model specification

$$Y_{ij}^{post} = \beta_0 + \beta_1 Y_{ij}^{pre} + \beta_2 T_{2ij} + \beta_3 T_{3ij} + u_j + \epsilon_{ij}$$

where

- $Y_{ij}^{post}$ = posttest score for student *i* in school *j,*
- $Y_{ij}^{pre}$ = pretest (covariate),
- $T_{2ij}, T_{3ij}$ = dummy codes for Integrated and Segregated groups,
- $u_j$ = school-level random effect,
- $\epsilon_{ij}$ = teacher-level residual.

Appendix D

ANCOVA Results for SMAT Posttest Scores for grade 10 by Treatment Group (Controlling for Pretest Scores)

| Dependent Variable | Source | SS | df | MS | F | p | Partial $\eta^2$ |
|---|---|---|---|---|---|---|---|
| Algebra | Treatment Group | 876.23 | 2 | 438.12 | 14.21 | .000** | .174 |
| | Pretest (Covariate) | 592.87 | 1 | 592.87 | 19.23 | .000** | .181 |
| | Error | 2647.34 | 86 | 30.78 | | | |
| Geometry | Treatment Group | 1257.41 | 2 | 628.71 | 16.84 | .000** | .192 |
| | Pretest (Covariate) | 781.62 | 1 | 781.62 | 20.94 | .000** | .196 |
| | Error | 3208.16 | 86 | 37.30 | | | |
| Cultural Context | Treatment Group | 97.22 | 2 | 48.61 | 5.48 | .006** | .113 |
| | Pretest (Covariate) | 112.75 | 1 | 112.75 | 12.71 | .001** | .128 |
| | Error | 757.48 | 86 | 8.81 | | | |
| Total SMAT Score | Treatment Group | 7638.51 | 2 | 3819.25 | 18.72 | .000** | .218 |
| | Pretest (Covariate) | 6453.96 | 1 | 6453.96 | 31.63 | .000** | .269 |
| | Error | 17537.65 | 86 | 203.95 | | | |

All dependent variables represent student posttest SMAT domain scores, adjusted for corresponding pretest values.
Treatment groups include *Integrated* and *Segregated* (Control as reference).
ANCOVA adjusted for clustering at the school level (CR2 correction).
**p** values are two-tailed; ** indicates $p < .01$.

SS Stands for sum of squares; df for degree of freedom; ms mean squares

## Model Specification

$$Y_{ij}^{post} = \beta_0 + \beta_1 Y_{ij}^{pre} + \beta_2 T_{1ij} + \beta_3 T_{2ij} + u_j + \epsilon_{ij}$$

where

- $Y_{ij}^{post}$ = posttest score for student $i$ in school $j$
- $Y_{ij}^{pre}$ = pretest (covariate)
- $T_{1ij}, T_{2ij}$ = dummy codes for Integrated and Segregated groups
- $u_j$ = school-level random effect
- $\epsilon_{ij}$ = student-level residual

Appendix E
ANCOVA Results for Grade 12 SMAT Posttest Scores Adjusted for Pretest

| Dependent Variable | Source | SS) | df | MS | F | p | Partial $\eta^2$ |
|---|---|---|---|---|---|---|---|
| Total SMAT Score | Pretest (covariate) | 892.67 | 1 | 892.67 | 9.83 | .002 | .065 |
| | Treatment Group | 2449.32 | 2 | 1224.66 | 13.49 | .000 | .160 |
| | Error | 12896.12 | 142 | 90.78 | | | |
| Algebra | Pretest | 431.22 | 1 | 431.22 | 10.71 | .001 | .070 |
| | Treatment Group | 2108.85 | 2 | 1054.43 | 26.17 | .000 | .270 |
| | Error | 5719.33 | 142 | 40.27 | | | |
| Geometry | Pretest | 95.48 | 1 | 95.48 | 4.70 | .032 | .032 |
| | Treatment Group | 353.81 | 2 | 176.91 | 8.70 | .000 | .110 |
| | Error | 2895.26 | 142 | 20.39 | | | |
| Calculus | Pretest | 61.09 | 1 | 61.09 | 5.53 | .020 | .038 |
| | Treatment Group | 195.88 | 2 | 97.94 | 8.87 | .000 | .112 |
| | Error | 1567.15 | 142 | 11.03 | | | |
| Cultural Context | Pretest | 20.61 | 1 | 20.61 | 5.97 | .016 | .041 |
| | Treatment Group | 67.45 | 2 | 33.73 | 9.78 | .000 | .121 |
| | Error | 489.83 | 142 | 3.45 | | | |

**Note.** ANCOVA adjusted for baseline (pretest) scores. Standard errors are cluster-robust at the school level. Partial $\eta^2$ indicates effect size. *$p \leq .05$.

## Model Specification

$$Y_{ij}^{post} = \beta_0 + \beta_1 Y_{ij}^{pre} + \beta_2 T_{1ij} + \beta_3 T_{2ij} + u_j + \epsilon_{ij}$$

where

- $Y_{ij}^{post}$ = posttest score for student $i$ in school $j$
- $Y_{ij}^{pre}$ = pretest (covariate)
- $T_{1ij}, T_{2ij}$ = dummy codes for Integrated and Segregated groups
- $u_j$ = school-level random effect
- $\epsilon_{ij}$ = student-level residual