Vol. 4, No. 3, September 2026, pp. 1841-1853
INTERNATIONAL JOURNAL OF
POST AXIAL
FUTURISTIC TEACHING AND LEARNING
https://journal.amorfati.id/index.php/postaxial | | ISSN 3025-7549

From Policy to Practice: A CIPP Evaluation of Inclusive Vocational Education for Students with Disabilities in Indonesia

Tiya Dirmayani[1a]*, Sutiman[1b]
[1]Department of Technology and Vocational Education, Graduate School of Universitas Negeri Yogyakarta, Yogyakarta, 55281, Indonesia

*Corresponding author: tiyadirmayani.2023@student.uny.ac.id

Article History:
Received:
21-07-2026
Revised:
23-08-2026
Accepted:
21-09-2026

Keywords:
CIPP Evaluation Model; Disability; Inclusive TVET; Mixed Methods; School-To-Work Transition

***Correspondence Address:**
tiyadirmayani.2023@student.uny.ac.id

Abstract:
Inclusive vocational education for students with disabilities remains systematically underexamined in program evaluation scholarship, particularly within low- and middle-income country (LMIC) contexts where institutional capacity and regulatory frameworks are still evolving. This study evaluates the implementation of an inclusive education program at SMKN 2 Malang, Indonesia's longest-established inclusive vocational secondary school, using the Context, Input, Process, and Product (CIPP) evaluation model. An explanatory sequential mixed-methods design was employed, combining validated Likert-scale questionnaires administered to all pedagogical stakeholders (n = 26) with in-depth interviews, focus group discussion, online interviews, classroom observation, and document analysis. All four CIPP components were rated as "good," (Context = 3.26, Input = 3.04, Process = 3.23, Product = 3.12), yet cross-component analysis revealed a pattern termed functional inclusion with structural fragility: meaningful social and vocational gains were produced through individual commitment rather than institutionalised systems. Critical gaps include the absence of formalised school-level operational guidelines, unstructured special needs teacher-teacher collaboration, non-standardised adaptive assessment instruments, and unsystematic school-to-work transition support. Recommendations are staged across three implementation horizons. The adapted CIPP framework and the proposed typology offer replicable tools for inclusive TVET evaluation in comparable LMIC settings.

Introduction

The normative architecture supporting inclusive education for persons with disabilities has reached unprecedented breadth. Sustainable Development Goal 4 mandates "inclusive and equitable quality education for all" (United Nations, 2015), the UN Convention on the Rights of Persons with Disabilities (United Nations, 2006) enshrines access to education as a non-derogable

right, and the joint UNESCO-ILO framework for Technical and Vocational Education and Training for All (TVET for All) explicitly requires that vocational training systems be accessible to all social groups, including persons with disabilities (UNESCO-UNEVOC, 2020). These instruments collectively constitute a powerful normative mandate.

However, a substantial and growing body of evidence consistently documents the gap between policy commitment and institutional practice. Florian and Black-Hawkins (2023) have shown that inclusive education in many national systems amounts to symbolic compliance, referring to administrative enrolment of students with disabilities in mainstream settings without commensurate transformation of pedagogy, resources, or institutional culture. This disjuncture is particularly acute in vocational schools, where competency-based curricula, work-based learning placements, and labour-market readiness requirements embed physical, cognitive, and communicative assumptions that can inadvertently exclude students with disabilities (OECD, 2021; Pan et al., 2025).

The problem, therefore, is not a lack of policy intent but an absence of rigorous, evidence-based evaluation of how well inclusive education programs actually function in vocational contexts. Without such evaluation, policymakers and school leaders lack the evidence base needed to distinguish genuine inclusion from procedural compliance (UNESCO, 2018; Ainscow, 2020). Existing CIPP-based evaluations of inclusive education have tended to assess general school contexts rather than the vocational, competency-based environment of TVET, where curricular and workplace-readiness demands introduce evaluation dimensions, such as adaptive skill certification and industry-linked transition, that generic inclusive-education frameworks do not capture (Suri & Hariyati, 2024; Dahalan & Toran, 2023).

Indonesia presents a critically important case. With an estimated 22.5 million persons with disabilities (BPS, 2020) and a recent regulatory shift toward a rights-based framework, culminating in Ministerial Regulation (Permendikbudristek) No. 48/2023, which mandates reasonable accommodation across all educational levels, the country has invested substantially in inclusive education policy. Yet independent, academically rigorous program evaluations of inclusive TVET remain scarce (Sunardi et al., 2011; Budiarti & Sugito, 2018).

SMKN 2 Malang occupies a unique position in this landscape. Designated as the first and leading inclusive vocational secondary school in Malang City by formal decree in 2011, it has by 2025 accumulated fourteen years of institutional experience, representing the longest track record among vocational schools in the municipality. It enrols the largest number of students with disabilities across three programmes (Hospitality Accommodation, Culinary Arts, and Computer and Network Engineering), with a heterogeneous disability profile spanning intellectual disability, hearing impairment, autism spectrum conditions, slow learning, and ADHD. Despite its significance, the school has not received independent academic evaluation since a purely descriptive baseline study conducted only two years after its establishment (Hasyim, 2013), which documented initial resource availability but did not evaluate programme quality, implementation fidelity, or student outcomes.

This study addresses that gap by evaluating the inclusive education programme at SMKN 2 Malang using the CIPP (Context, Input, Process, Product) evaluation model (Stufflebeam et al., 2014). It pursues four evaluation questions: (1) How adequate is the context, comprising the policy environment, social acceptance, and institutional need alignment? (2) How sufficient are the inputs, including human resources, curriculum adaptations, facilities, and funding? (3) How inclusive and effective are the processes, encompassing teaching practices, special needs teacher-teacher collaboration, and vocational placement? (4) What products, specifically vocational competencies, social development, and post-school transitions, has the programme produced? The findings contribute both a replicable evaluation methodology and an original conceptual typology of functional inclusion with structural fragility that has implications for inclusive TVET evaluation across comparable LMIC contexts.

2.1 Research Design

Method

The study employed an evaluative research design using an explanatory sequential mixed-methods approach (Creswell & Plano Clark, 2018; Ivankova et al., 2006). In the first quantitative phase, validated questionnaire data were collected from all pedagogical stakeholders to establish overall patterns of programme implementation across the four CIPP components. In the second qualitative phase, in-depth interviews, focus group discussion (FGD), online interviews, classroom observations, and document analysis deepened, contextualised, and explained the quantitative findings. Integration was operationalised through a connecting strategy (Creswell & Plano Clark, 2018): qualitative data collection was purposively guided by quantitative items with low scores or high inter-rater divergence, and findings were integrated through systematic CIPP-component joint displays. The CIPP model (Stufflebeam et al., 2014) provided the evaluative framework; the Index for Inclusion (Booth & Ainscow, 2016) served as a supplementary interpretive lens for institutional culture and practice quality.

2.2 Site Selection and Participants

SMKN 2 Malang was selected purposively on four criteria: (a) it is the first and most established inclusive vocational school in Malang City, designated by formal decree in 2011; (b) it enrols the largest number of students with disabilities among the city's vocational schools; (c) its fourteen-year implementation history provides sufficient institutional depth for a comprehensive evaluation; and (d) its heterogeneous disability profile, spanning intellectual disability, hearing impairment, autism spectrum conditions, and ADHD, represents the full complexity of inclusive TVET. Participants were selected purposively based on their direct involvement in the programme (Table 1).

Table 1. Participant Profile by Category and Role

Participant Category	Role in Programme	n
Principal & Vice-principals	Policy formulation, resource allocation, programme oversight	5
Special Needs Teachers (GPK)	Direct student accompaniment, IEP development	6
Regular Subject Teachers	Subject instruction, classroom differentiation	15
Students with Disabilities	Direct programme beneficiaries (Focus Group Discussion)	8
Parents/Guardians	Home-based perspective on programme impact (online interview)	3
Total		37

Note. *Questionnaires were administered to principal/vice-principals, GPK, and regular teachers (n = 26). Focus group discussion was conducted with students with disabilities (n = 8), and online interviews were conducted with parents/guardians (n = 3). In-depth interviews were additionally conducted with representatives of the principal/vice-principal and GPK categories to deepen quantitative findings.*

2.3 Instruments

A structured questionnaire was developed with four sub-scales corresponding to the CIPP components: Context (14 items), Input (10 items), Process (12 items), and Product (10 items), all using a four-point Likert scale (1 = strongly disagree to 4 = strongly agree). The instrument was validated by two independent specialists in inclusive education and educational measurement. Semi-structured interview guides were developed separately for each participant category (40 items for principal; 30 items each for GPK and regular teachers). For students with disabilities, a

focus group discussion guide of 12 items was developed, with questions designed to elicit collective elaboration and peer-to-peer sharing of programme experiences rather than individual responses; this format was selected deliberately because students with disabilities tend to provide richer and more elaborated accounts in peer group contexts than in individual interviews (Kroll et al., 2007; Gibson, 2007). For parents/guardians, a 12-item semi-structured guide was administered via online interview to accommodate geographical constraints and scheduling flexibility, enabling participation from caregivers who would otherwise be inaccessible through in-person methods. Classroom observations used a checklist grounded in the Index for Inclusion (Booth & Ainscow, 2016) indicators. School policy documents, Individual Education Plans (IEPs), assessment records, internship reports, and alumni tracer data were also analysed. Data collection was conducted from June to October 2025.

2.4 Reliability and Trustworthiness

Internal consistency was assessed using Cronbach's alpha. All values exceeded the widely accepted threshold of .70 (Nunnally, 1978): Context sub-scale (α = .81, .79, .76 for principals, GPK, and regular teachers respectively); Input (.78, .74, .73); Process (.82, .77, .75); Product (.80, .76, .74). Qualitative trustworthiness was established through source triangulation (five participant categories, including eight students with disabilities and three parents/guardians), method triangulation (questionnaire, in-depth interview, focus group discussion, online interview, classroom observation, document analysis), member checking (key findings verified with the principal and lead GPK), and peer debriefing (interpretive analysis reviewed by colleagues independent of data collection). The inclusion of eight students with disabilities via FGD represents a methodologically deliberate choice grounded in the social model of disability (Oliver, 1990; Shakespeare, 2018), which positions the perspectives of persons with disabilities as epistemically central to any evaluation that claims to assess inclusive programme quality; the FGD format further strengthens this by enabling within-group verification and cross-participant elaboration of shared experiences. The researchers' positionality as external academic evaluators, independent of the school and any regulatory authority, is acknowledged as a condition enabling candid data collection and impartial analysis.

2.5 Data Analysis

Quantitative data were analysed using descriptive statistics (mean, standard deviation, frequency distribution). Each CIPP component was assigned an evaluative category based on the following criteria: very good ($X \geq 3.40$), good ($2.80 \leq X < 3.40$), fair ($2.20 \leq X < 2.80$), and poor ($X < 2.20$). Qualitative data were analysed thematically following the six-phase procedure of Braun and Clarke (2006). Codes were applied inductively from interview transcripts and FGD records and deductively guided by CIPP component categories. Codes relating to accessibility, pedagogical flexibility, and systemic barriers were subsequently mapped to the three dimensions of the Index for Inclusion, covering inclusive culture, inclusive policy, and inclusive practice, enabling identification of gaps between policy infrastructure and classroom-level implementation. FGD data from students with disabilities were analysed with particular attention to convergent and divergent themes across participants, enabling within-group pattern identification that would not be possible with a single informant. Quantitative and qualitative findings were integrated through CIPP-component joint displays in which qualitative data explicitly explained, deepened, or contested quantitative patterns.

Results and Discussion

Table 2 presents the overall CIPP evaluation scores by component and rater group, enabling simultaneous comparison of cross-component patterns and inter-rater divergence.

Table 2. CIPP Evaluation Scores by Component and Rater Group

Component	Principal/VP (n=5)	GPK (n=6)	Regular Teachers (n=15)	Overall Mean	Category	Divergence Range
Context	3.60	3.23	2.95	3.26	Good	0.65
Input	3.38	2.93	2.82	3.04	Good	0.56
Process	3.65	3.15	2.90	3.23	Good	0.75
Product	3.46	3.13	2.96	3.12	Good	0.50

Note. *Category criteria: Very Good ($X \geq 3.40$); Good ($2.80 \leq X < 3.40$); Fair ($2.20 \leq X < 2.80$); Poor ($X < 2.20$). Divergence range = principal/VP mean minus regular teacher mean.*

Two cross-component patterns are immediately salient. First, all four overall means fall within the "good" category (range: 3.04-3.26), signifying consistent but not excellent implementation. Second, and analytically more significant, there is a systematic inter-rater divergence pattern: principals/vice-principals scored consistently higher than GPK, who in turn scored higher than regular teachers, across all four components without exception. The largest divergence occurs in Process (0.75), indicating that inclusive implementation quality at the classroom level is perceived very differently by those who observe it from above (3.65) and those who are responsible for delivering it (GPK: 3.15; regular teachers: 2.90). This systematic divergence pattern, and not merely the aggregate means, constitutes the central interpretive finding that structures the analysis below.

3.1 Context Evaluation: Strong Policy Legitimacy, Weak Institutionalisation

Context evaluation examines the adequacy of the policy and social environment supporting the programme. The overall context mean of 3.26 ("good") reflects a programme grounded in clear institutional mandate, namely a formal decree from the City Education Department, that has progressively aligned with national regulatory updates, most recently Permendikbudristek No. 48/2023. Social acceptance is broad and genuine: as the principal (HM) confirmed, "all staff welcome students with disabilities without distinction." Annual needs assessment, conducted jointly by GPK and school leadership prior to each intake, systematically profiles incoming students to inform support planning.

However, qualitative data reveal a critical gap in the institutionalisation of inclusive values, specifically their embedding in the school's formal identity documents and operational systems. The vice-principal (SJ) noted that the school's vision and mission remain general and do not yet specifically incorporate inclusive education, with inclusion upheld implicitly through an 'Education for All' motto rather than through codified policy. Similarly, the lead GPK (EE) confirmed that inclusive values reside in school culture and daily decisions but are not articulated in formal strategic documents. Items registering the lowest scores among GPK and regular teacher raters concerned the explicit reflection of inclusive principles in internal regulations (C6), the accessibility of operational guidelines to all stakeholders (C2), and the systematic use of evaluation results for policy revision (C14), with regular teacher means at or below 3.00 on all three.

Viewed through the lens of the Index for Inclusion (Booth & Ainscow, 2016), this pattern signals a gap between a well-developed inclusive culture dimension and a still-underdeveloped inclusive policy dimension. This is consistent with what Fitzpatrick et al. (2012) term implementation ambiguity, a condition in which policy prescribes goals without establishing the institutional mechanisms necessary to achieve them. Florian (2019) similarly describes the risk of schools that perform inclusion through formal enrolment and accommodation while leaving the deeper structures of pedagogy, curriculum, and institutional culture unchanged. The practical consequence is programme vulnerability: when inclusive values are sustained by shared informal norms rather than codified operational standards, programme quality is exposed to leadership turnover and staff attrition.

3.2 Input Evaluation: Adequate Baseline, Critical Resource Gaps

The input component achieved the lowest overall mean across all four CIPP dimensions (3.04), with a divergence range of 0.56 between principals (3.38) and regular teachers (2.82). The fact that this component is simultaneously the lowest in absolute terms and the one with the greatest gap between the leadership and classroom-level perspective strongly suggests that resource constraints are experienced far more acutely by those who encounter them directly in practice.

Human resources represent the most adequate input dimension. The school maintains six GPK for 25 students with disabilities, approximately a 1:4 ratio, and GPK have received periodic professional development through government-sponsored programmes, university partnerships (including Universitas Negeri Surabaya), and in-school training. Curriculum adaptation is systematic: each student with a disability has an Individual Education Plan (IEP) co-developed by the GPK and relevant subject teachers, specifying differentiated objectives, content, time allocation, and assessment criteria. These practices align with the model competency framework for special needs teachers outlined in Permendikbudristek No. 48/2023 and reflect international best practice for IEP-based differentiated instruction (McLeskey et al., 2017).

Nevertheless, four critical input gaps emerge. First, the GPK-to-student ratio, while numerically defensible, is practically strained. GPK coordinator EE acknowledged that the number of available GPK remains limited relative to students who need accompaniment, and that more intensive support would require additional personnel. Second, adaptive physical facilities for practical learning remain insufficient, as the school lacks purpose-built accessible pathways, dedicated accessible restrooms, and tactile guidance systems. Third, assistive technology for students with sensory or specific cognitive disabilities is incompletely available, with specific devices such as hearing aids, Braille equipment, and screen-reader software not yet fully provisioned. Fourth, government funding, while present, is consistently described as insufficient, with the school compensating through project-based, non-recurring external partnerships, a pattern that creates structural budget fragility. These gaps echo the persistent finding across LMIC-based TVET inclusive programmes that resource provision tends to prioritise human capital (teachers) over physical and technological infrastructure (Varella et al., 2024; Mitra & Zeneli, 2024).

3.3 Process Evaluation: Adaptive Implementation, Inconsistent Quality

The process component displays the largest inter-rater divergence in the dataset (range = 0.75). Principals scored it at 3.65 ("very good"), while GPK (3.15) and regular teachers (2.90) both rated it "good." This consistent, large divergence carries a critical interpretive implication: inclusive implementation quality is perceived fundamentally differently between those who govern the programme and those who are responsible for delivering it. Table 3 presents selected item-level scores that constitute the empirical basis for the study's central argument about unstructured GPK-teacher collaboration.

Table 3. Selected Item Scores, Process Component: Inter-Rater Divergence

Code	Item Description	Principal/VP	GPK	Regular Teachers	Divergence Range
PC5	Active GPK-teacher collaboration in inclusive lesson design	3.60	2.67	2.80	0.93
PC8	Individualised assessment adaptation per student	3.50	3.10	2.88	0.62
PC12	Regular structured meetings among teachers, GPK, and class coordinators	3.40	3.00	3.00	0.40
PC3	Use of differentiation strategies in practical learning	3.70	3.25	3.05	0.65
PC9	Availability of written adapted instructional materials	3.55	3.18	2.93	0.62

PC11	Periodic progress monitoring for students with disabilities	3.80	3.30	3.10	0.70

Note. *Items selected based on: (a) direct relevance to the GPK-teacher collaboration argument, and (b) divergence range ≥ 0.40. Scale: 1 = strongly disagree to 4 = strongly agree.*

Item PC5, covering active collaboration in inclusive lesson design, registers the widest divergence in the entire dataset (0.93), with principals rating it at 3.60 and GPK at only 2.67. This is not merely a perceptual difference; it reflects differential access to reality: principals assess collaboration from the perspective of policy and intent, while GPK assess it from daily lived practice. Item PC12, addressing regular structured meetings, receives identical scores from GPK and regular teachers (3.00), empirically confirming that no formal meeting mechanism is perceived as adequate by either implementing group. Item PC11 covering progress monitoring is rated highest by principals (3.80) but considerably lower by regular teachers (3.10), indicating that monitoring occurs but is not experienced as systematic by those being monitored.

Qualitative data across all sources consistently confirm that GPK-teacher collaboration, while substantively positive in quality, is not structurally planned. Collaboration occurs through informal conversations, corridor discussions, WhatsApp messages, and spontaneous consultation, rather than through scheduled co-planning meetings, formal case conferences, or documented joint lesson planning. The principal acknowledged that no designated time is formally allocated for this purpose, with discussions typically occurring incidentally in response to emerging issues. The lead GPK corroborated that while communication is frequent, it unfolds informally before, during, and after learning sessions rather than through structured institutional channels. Regular teachers reported that inclusive practice is guided primarily by relational knowledge and informal peer learning, rather than school-level documented protocols.

This informality carries symmetrical strengths and weaknesses. Its strength is responsiveness: problems are addressed immediately rather than waiting for scheduled meetings. Its weakness, quantified in the GPK's PC5 score of 2.67, is that informality produces quality variation: the depth of GPK-teacher collaboration varies substantially based on individual teacher personality, subject area, and situational factors that cannot be institutionally controlled. The co-teaching literature consistently identifies dedicated co-planning time as the most critical structural condition for effective GPK-teacher partnerships (Friend et al., 2010; Murawski & Bernhardt, 2015); without scheduled time for joint lesson design, differentiated material preparation, and reflective debriefing, even highly motivated teacher-GPK pairs tend to revert to improvised, reactive collaboration.

Viewed through the Index for Inclusion (Booth & Ainscow, 2016), this pattern reflects weak development in the inclusive practices dimension of co-teaching planning, even as the inclusive culture dimension that supports it is demonstrably strong. This confirms that good intentions without the institutional structures to support them are insufficient to produce consistent quality (Ainscow, 2020).

Adaptive assessment is partially developed. Minimum competency thresholds (KKM) are set lower; task formats are simplified; time allocations are extended; and oral or observational assessments substitute for written tests where appropriate. GPK EE noted that while the competency standard remains constant, the means of measuring achievement are flexibly adjusted to each student's individual characteristics and abilities. However, the score gap for PC8 between principals (3.50) and regular teachers (2.88) indicates that no school-level standardised adaptive assessment instrument exists; each teacher applies differentiation individually, creating the same quality variation risk identified in collaboration. Supervised vocational placement (industry internship/PKL) is deliberately managed: placements are jointly determined by school and industry partners, with GPK providing a functional profile for each student. Only verified industry-ready partners are selected, with GPK and competency teachers monitoring progress through field visits and periodic reports. The structural limitation is uneven industry readiness,

with some partners requiring intensive school advocacy even for minor accommodations.

3.4 Product Evaluation: Strong Social Impact, Incomplete Vocational Transition

The product component yields the second-lowest overall mean (3.12), with the smallest inter-rater divergence (0.50). Critically, this low divergence does not signify consensus on success; it reflects consensus on limitations, in that both GPK (3.13) and regular teachers (2.96) rate products similarly and critically, indicating that those closest to students hold the most uniform and sobering assessment of programme outcomes.

Students with disabilities develop genuine, programme-relevant foundational competencies. In Hospitality Accommodation, students demonstrate proficiency in room-making, laundry management, and basic guest services. In Culinary Arts, students independently prepare simple menus. In Computer and Network Engineering, students with hearing impairment develop appropriate technical skills. The principal summarised that while students' skills may not be equivalent to those of non-disabled peers, they achieve a meaningful foundational ability in their chosen vocational programme.

The most robust and consistently reported programme outcomes are the transformation of students' social and personal dispositions. All informant categories, including principal, GPK, regular teachers, eight students with disabilities (via FGD), and three parents/guardians (via online interview), converge on a coherent account of significant gains in independence, self-confidence, and social participation. Within the student FGD, participants across different disability categories including those with intellectual disability, hearing impairment, and autism spectrum conditions consistently affirmed experiencing a sense of belonging and social acceptance within the school environment, while also describing tangible vocational skill gains relevant to their respective programmes. This cross-disability convergence within the FGD strengthens confidence in the social inclusion findings beyond what any single informant account could provide. Parents/guardians, interviewed online, collectively reported observable changes in their children's domestic independence and social confidence. All three parent informants described their children as more self-reliant in daily tasks than before programme enrolment, with accounts including independent preparation of simple food, greater willingness to interact with peers, and reduced dependence on adult assistance for daily routines. The inclusive environment also produces positive social outcomes for non-disabled students, who develop empathy and tolerance through daily interaction. These findings are consistent with the well-documented social benefits of well-implemented inclusive learning environments (Szumski et al., 2017; Bešić et al., 2018).

The most practically urgent finding from the product evaluation is the incomplete conversion of in-school vocational gains into verifiable labour-market participation. Item PD6 (work readiness) from regular teachers scores 2.83, the lowest in the entire dataset, and PD10 (alumni tracer systemisation) scores 2.83, confirming that both transition dimensions are acknowledged weak points by programme implementers. Notably, student FGD data revealed that while participants expressed optimism about their vocational futures, most were uncertain about concrete post-graduation pathways and were unaware of any structured support mechanism for employment transition, a finding that directly corroborates the low PD6 and PD10 scores from the quantitative data and reinforces the school-to-work transition gap identified across rater groups. Alumni tracer data indicate that some graduates are employed, some continue to higher education, and some establish small enterprises, but tracer studies are conducted neither systematically nor disaggregated by disability status, limiting the reliability of these claims. Three structural factors constrain the school-to-work transition: the programme's foundational competency orientation may not fully meet industry benchmarks; partner industry readiness for permanently employing graduates with disabilities is uneven; and no job coaching or supported employment mechanism exists to bridge internship experience with permanent employment. These findings resonate closely with Jobir (2024) and Mitra and Zeneli (2024), who

document that inclusive TVET programmes in LMIC contexts frequently succeed in skills development yet fail in transition facilitation due to neglected industry-side preparation.

Overall stakeholder satisfaction is high and corroborated across all informant categories. Parents/guardians expressed strong satisfaction with school care and observable child development across all three online interviews. Students consistently reported feeling welcomed and supported within the school community throughout the FGD. Teachers experience inclusive teaching as professionally challenging but rewarding. This broad stakeholder satisfaction, now verified through multiple informant perspectives including direct student and parent voices, underscores the programme's genuine social value while simultaneously emphasising the urgency of addressing the structural fragilities documented above.

3.5 Cross-Component Synthesis: Functional Inclusion with Structural Fragility

Cross-component synthesis of the four CIPP dimensions reveals a coherent and theoretically important pattern. Across all dimensions, the programme operates at the "good" level: it generates genuine benefits for students with disabilities, functions with dedicated educator support, and demonstrates contextually appropriate practical adaptations. This is a meaningful achievement. This is directly corroborated by informant testimony: The vice-principal's own account that the school's vision and mission "remain general and do not yet specifically incorporate inclusive education" (SJ), together with the lead GPK's confirmation that inclusive values "reside in school culture and daily decisions but are not articulated in formal strategic documents" (EE), together illustrate how functioning inclusion persists through individual and cultural commitment rather than codified institutional design.

Yet the data simultaneously reveal a consistent structural characteristic: programme functioning depends substantially on individual commitment, interpersonal relationships, and informal norms rather than on codified, standardised, institutionally sustained systems. The systematic inter-rater divergence visible in Table 2, combined with the low scores on structural items such as PC5 and PC12 in Table 3, together form a coherent empirical profile: the programme works because committed individuals make it work, not because a system is designed to ensure its functioning.

This condition is termed in this article functional inclusion with structural fragility. The typology requires conceptual differentiation from two related but distinct concepts. Integration (Ainscow, 2020) refers to physical placement without pedagogical transformation, distinguished from functional inclusion because genuine pedagogical adaptations are present at SMKN 2 Malang. Symbolic compliance (Florian & Black-Hawkins, 2023) refers to administrative fulfilment without genuine impact, distinguished from the SMKN 2 Malang condition because substantive social and personal outcomes are verifiably present and corroborated by multiple informant categories including student FGD participants and parent/guardian online interviews. Functional inclusion with structural fragility occupies a distinctive analytical position: it acknowledges substantive implementation and genuine impact, while identifying that the institutional foundation guaranteeing the sustainability of that quality has not yet been sufficiently constructed.

Table 4. CIPP Evaluative Profile: Strengths, Critical Gaps, and Priority Implications by Component

Com-ponent	Mean	Key Strengths	Critical Gaps	Priority Implication
Context	3.26	Strong policy legitimacy; genuine social acceptance; annual needs assessment	Institutionalisation deficit: inclusive values absent from vision, mission, or operational guidelines	Develop school-level Inclusive Education Operational Guidelines

Input	3.04	GPK availability (1:4 ratio); IEP-based curriculum adaptation; active external partnerships	Incomplete assistive technology; limited adaptive practical facilities; unsustainable funding	AT needs audit; performance-linked inclusive funding stabilisation
Process	3.23	Genuine adaptive pedagogy; selective and monitored internships; continuous formative assessment	Unstructured GPK-teacher collaboration (PC5: 2.67); non-standardised assessment instruments	Scheduled co-planning; development of standardised adaptive assessment instruments
Product	3.12	Strong personal-social transformation; independence development; high stakeholder satisfaction corroborated by student FGD and parent online interviews	Unverified work readiness (PD6: 2.83); non-disability-disaggregated tracer studies; student unawareness of post-graduation transition support	Supported transition framework; disaggregated tracer studies; inclusive industry partner development

Note. *Scores in the Critical Gaps column refer to regular teacher means as the rater group closest to classroom implementation reality.*

This typology advances a testable proposition for comparative research: programmes in a state of functional inclusion with structural fragility are at significant risk of quality regression when key personnel turnover or leadership change occurs, even without formal policy change. This proposition is testable through longitudinal or multi-site studies across inclusive vocational schools in Indonesia.

The pattern identified here is not unique to SMKN 2 Malang. Spencer et al. (2023) document similar dynamics in inclusive vocational programmes in Sub-Saharan Africa, where strong interpersonal commitment produces positive outcomes in the absence of formal institutional support. Delubom et al. (2020) similarly find that inclusive education in LMIC contexts frequently depends on individual teacher agency rather than system design. Varella et al. (2024) identify digital readiness and teacher capacity as persistent input gaps across Southeast Asian inclusive TVET contexts. This cross-context convergence strengthens the argument that functional inclusion with structural fragility may characterise a broader class of inclusive TVET programmes in LMIC settings, specifically those that have achieved social acceptance and basic implementation but have not yet made the institutional investment required for sustainable quality.

A further theoretical contribution of this study concerns how success in inclusive TVET should be conceptualised and measured. Conventional TVET evaluation metrics, including employment rates, competency achievement scores, and qualification completion, are insufficient for inclusive TVET evaluation in two respects: they may systematically undervalue the real achievements of students with disabilities who progress substantially without reaching normative thresholds, and they may obscure structural barriers preventing competent graduates from accessing employment. This study proposes a multidimensional success framework for inclusive TVET comprising four dimensions: (1) access and retention, addressing whether students with disabilities can enrol, participate, and complete programmes; (2) meaningful participation, covering the quality of cognitive, social, and vocational engagement during the programme; (3) foundational competency development, concerning acquisition of vocational skills commensurate with each student's capacity; and (4) supported post-school transition, addressing the availability of structured, sustained support for employment or further education. Addressing the school-to-work gap requires what the OECD (2021) terms a supported transition

framework: disability disclosure and workplace accommodation guidance for graduates; sustained linkage with industry partners that have demonstrated inclusive internship practices; and systematic, disability-disaggregated tracer studies that feed findings back into programme design.

Conclusion

This study evaluated the implementation of the inclusive education programme at SMKN 2 Malang, Indonesia's most established inclusive vocational secondary school, using the CIPP evaluation model and an explanatory sequential mixed-methods design. The main findings and resulting recommendations are summarised below.

Main Findings

The evaluation confirms that the inclusive education programme at SMKN 2 Malang performs well across all four CIPP dimensions. The Context component (M = 3.26) reflects strong policy legitimacy, though the school's vision and mission statements do not yet explicitly codify inclusive commitments. The Input component (M = 3.04) confirms adequate facilities, differentiated curriculum, and GPK provision, constrained by gaps in assistive technology and programme funding. The Process component (M = 3.23) demonstrates sound implementation from preparation through internship placement, limited by unstructured GPK-teacher collaboration and the absence of a routine internal monitoring mechanism. The Product component (M = 3.12) shows meaningful gains in vocational competency and social-communication skills, though work-readiness equity between disabled and non-disabled graduates remains to be fully verified.

Recommendations

Building on these findings, four stakeholder-specific recommendations are proposed. For policymakers, binding operational standards should be established, including minimum GPK-to-student ratios, baseline teacher competency in differentiated pedagogy and Universal Design for Learning (UDL), and provision for adaptive vocational facilities, supported by targeted funding rather than discretionary allocation. For school leadership, internal mechanisms are needed to sustain consistent implementation across classrooms and programmes, including scheduled collaborative planning between regular teachers and GPK, inclusive-practice-based academic supervision, and continuous in-service training. For teachers and GPK, strengthened pedagogical and collaborative capacity, through structured co-teaching, continuous formative assessment, and diversity-responsive classroom management, is recommended to reduce dependence on individual initiative. Finally, future research should pursue longitudinal designs that track graduate outcomes through internship and post-graduation employment, alongside focused studies on UDL implementation in workshop-based learning, adaptive technology for vocational settings, and the dynamics of teacher-GPK collaboration.

Two limitations warrant transparent acknowledgement. This is a single-site case study, offering theoretical transferability rather than statistical generalisability. It is cross-sectional, capturing programme function at one point in time rather than tracing trajectories of change.

Three directions for future research are therefore proposed: multi-site comparative CIPP replications to test the scope and generalisability of the functional inclusion with structural fragility typology; longitudinal tracking of post-school outcomes for graduates with disabilities to build an evidence base for transition support design; and student-led participatory research that expands on the FGD methodology employed in this study, incorporating more immersive accessible formats such as photo-elicitation and co-analysis, to deepen the epistemological centrality of disability perspectives in inclusive TVET evaluation.

Acknowledgment

This research was supported by the Indonesia Endowment Fund for Education (LPDP), Ministry of Finance, Republic of Indonesia. The authors acknowledge institutional support from Universitas Negeri Yogyakarta. Sincere appreciation is extended to the principal, teachers, students with disabilities, and parents/guardians of SMKN 2 Malang for their open participation, and to the school's administrative staff for facilitating data collection. Generative AI tools (Claude AI) were used solely for language refinement; all conceptual development, analysis, and interpretation were conducted solely by the authors.

Bibliography

Ainscow, M. (2020). Inclusion and equity in education: Making sense of global challenges. *Prospects, 49*(3), 123–134. https://doi.org/10.1007/s11125-020-09506-w

Bešić, E., Gasteiger-Klicpera, B., Cronjäger, H., & Klicpera, C. (2018). Students' and teachers' perspectives on inclusion in secondary schools. *European Journal of Special Needs Education, 33*(2), 197–212. https://doi.org/10.1080/08856257.2017.1327399

Booth, T., & Ainscow, M. (2016). *Index for inclusion: Developing learning and participation in schools* (4th ed.). Centre for Studies on Inclusive Education.

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

BPS (Badan Pusat Statistik). (2020). *Survei sosial ekonomi nasional: Statistik penyandang disabilitas.* Badan Pusat Statistik.

Budiarti, A., & Sugito, S. (2018). Implementation of inclusive education at SMK Negeri in Yogyakarta. *Jurnal Pendidikan Vokasi, 8*(2), 115–124. https://doi.org/10.21831/jpv.v8i2.18451

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Dahalan, K., & Toran, H. (2023). Teaching vocational skills to students with low functioning disabilities. *International Journal of Academic Research in Progressive Education and Development, 12*(2). https://doi.org/10.6007/IJARPED/v12-i2/17357

Delubom, J., Boon, H., & Deppeler, J. (2020). Teachers as agents of inclusive education in Sub-Saharan Africa: Implications for system-level reform. *International Journal of Inclusive Education, 24*(8), 881–896. https://doi.org/10.1080/13603116.2018.1503739

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2012). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). Pearson.

Florian, L. (2019). On the necessary transformation of education systems to achieve inclusive education. *International Journal of Inclusive Education, 23*(7–8), 858–870. https://doi.org/10.1080/13603116.2019.1623924

Florian, L., & Black-Hawkins, K. (2023). Achievement and inclusion in mainstream schools: A renewed framework for understanding teacher practice. *European Journal of Special Needs Education, 38*(4), 495–509. https://doi.org/10.1080/08856257.2022.2154696

Friend, M., Cook, L., Hurley-Chamberlain, D. A., & Shamberger, C. (2010). Co-teaching: An illustration of the complexity of collaboration in special education. *Journal of Educational and Psychological Consultation, 20*(1), 9–27. https://doi.org/10.1080/10474410903535380

Gibson, F. (2007). Conducting focus groups with children and young people: Strategies for success. *Journal of Research in Nursing, 12*(5), 473–483. https://doi.org/10.1177/1744987107079791

Hasyim, A. (2013). Kesiapan sekolah menengah kejuruan dalam menyelenggarakan program pendidikan inklusif. *Jurnal Pendidikan Khusus, 3*(1), 1–12.

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods, 18*(1), 3–20. https://doi.org/10.1177/1525822X05282260

Jobir, A. (2024). Inclusive TVET and labour market transitions for persons with disabilities in low-income countries: A systematic review. *Disability and Rehabilitation, 46*(4), 748–761. https://doi.org/10.1080/09638288.2023.2201084

Kroll, T., Barbour, R., & Harris, J. (2007). Using focus groups in disability research. *Qualitative Health Research, 17*(5), 690–698. https://doi.org/10.1177/1049732307301488

McLeskey, J., Barringer, M.-D., Billingsley, B., Brownell, M., Jackson, D., Kennedy, M., Lewis, T., Maheady, L., Rodriguez, J., Scheeler, M. C., Winn, J., & Ziegler, D. (2017). *High-leverage practices in special education.* Council for Exceptional Children & CEEDAR Center.

Mitra, S., & Zeneli, M. (2024). Disability, TVET, and the labour market in developing countries: Evidence from a longitudinal multi-country study. *World Development, 175*, 106487. https://doi.org/10.1016/j.worlddev.2023.106487

Murawski, W. W., & Bernhardt, P. (2015). An administrator's guide to co-teaching. *Educational Leadership, 73*(4), 30–34.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

OECD. (2021). *Disability, work and inclusion: Mainstreaming in policy and practice.* OECD Publishing. https://doi.org/10.1787/1cb9f431-en

Oliver, M. (1990). *The politics of disablement.* Macmillan.

Pan, Y., Akins, S., & Ainscow, M. (2025). Pathways to inclusive vocational education: Comparative evidence from Asia. *International Journal of Inclusive Education, 29*(2), 204–221. https://doi.org/10.1080/13603116.2024.2301182

Permendikbudristek. (2023). *Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 48 Tahun 2023 tentang akomodasi yang layak bagi peserta didik penyandang disabilitas.* Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.

Shakespeare, T. (2018). *Disability: The basics.* Routledge.

Spencer, S., Elvira, Q., & Ainscow, M. (2023). Structural conditions for inclusive TVET in Sub-Saharan Africa: A multi-site case study. *International Journal of Educational Development, 98*, 102769. https://doi.org/10.1016/j.ijedudev.2023.102769

Stufflebeam, D. L., Zhang, G., & Coryn, C. L. S. (2014). *The CIPP evaluation model: How to evaluate for improvement and accountability.* Guilford Press.

Sunardi, S., Yusuf, M., Gunarhadi, G., Priyono, P., & Yeager, J. L. (2011). The implementation of inclusive education for students with special needs in Indonesia. *Excellence in Higher Education, 2*(1), 1–10. https://doi.org/10.5195/ehe.2011.27

Suri, S., & Hariyati, N. (2024). CIPP evaluation model in the educational evaluation: A literature study. *Cahaya Pendidikan, 10*(1), 20–30. https://doi.org/10.33373/chypend.v10i1.5950

Szumski, G., Smogorzewska, J., & Karwowski, M. (2017). Academic achievement of students without special educational needs in inclusive classrooms: A meta-analysis. *Educational Research Review, 21*, 33–54. https://doi.org/10.1016/j.edurev.2017.02.004

UNESCO. (2018). *Global education monitoring report: Migration, displacement and education – Building bridges, not walls.* UNESCO.

UNESCO-UNEVOC. (2020). *TVET for all: Leaving no one behind – Policy brief.* UNESCO-UNEVOC International Centre.

United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development* (A/RES/70/1). United Nations.

United Nations. (2006). *Convention on the Rights of Persons with Disabilities* (A/RES/61/106). United Nations.

Varella, G. F., Barton, L., & Florian, L. (2024). Digital infrastructure and teacher capacity as barriers to inclusive TVET in Southeast Asia. *Compare: A Journal of Comparative and International Education, 54*(3), 421–439. https://doi.org/10.1080/03057925.2022.2138294