Asia Pacific Academy of Science Pte. Ltd.

Sustainable Social Development 2026, 4(5), 8992.
https://doi.org/10.23812/ssd8992

Article

Promoting SDG 4 through CIPP-based evaluation of vocational education training bases

Xiaoyi Shu, Premsuree Chaumthong*

Institute of Science, Innovation and Culture (ISIC), Rajamangala University of Technology Krungthep, Bangkok 10120, Thailand
*** Corresponding author:** Premsuree Chaumthong, premsuree.c@mail.rmutk.ac.th

CITATION

Shu X, Chaumthong P. Promoting SDG 4 through CIPP-based evaluation of vocational education training bases. Sustainable Social Development. 2026; 4(5): 8992. https://doi.org/10.23812/ssd8992

ARTICLE INFO

Received: 20 July 2026
Revised: 15 August 2026
Accepted: 26 August 2026
Available online: 14 September 2026

Abstract: Vocational training bases connect classroom learning with industry practice, but provincial quality reports rarely distinguish the system conditions, resources, implementation processes, and outcomes that shape their contribution to Sustainable Development Goal 4 (SDG 4). This study used a descriptive evaluative case-study design, combining documentary analysis with secondary analysis of the 2023 Guangdong Provincial Higher Vocational Education Quality Report. The unit of analysis was Guangdong's provincial higher-vocational education system; training-base performance was examined through indicators directly related to practical capacity, financing, staffing, industry participation, curriculum delivery, certification, graduate transitions, and stakeholder outcomes. Of 123 appendix indicators, 121 contained values, and 40 met prespecified relevance, traceability, non-duplication, and interpretability criteria. Absolute, relative, and direction-adjusted changes were calculated for comparable 2022–023 indicators, while conflicting source values were retained rather than averaged. The 2023 profile documented 450 virtual-simulation bases, 216 enterprise co-built practice centers 853 teaching-resource repositories, 19,359 industry mentors, and 109,182 vocational-skill certificate recipients. In the primary trend set, Process indicators had the largest median direction-adjusted change (18.48 %), followed by Input (10.07 %) and Product (1.52 %). The graduate destination rate fell from 98.62 % to 92.16 %, and fiscal appropriation per student declined by 13.23 %. These descriptive results are consistent with a possible mismatch between faster growth in selected capacity and process indicators and slower movement in several outcomes; they do not establish causation. The study provides an auditable CIPP-SDG 4 reporting procedure and identifies the need for utilization, equity, and longitudinal outcome data.

Keywords: quality assurance; industry–education integration; work-based learning; skills ecosystems; digital simulation; outcome monitoring; administrative data

1. Introduction

Technical and vocational education and training (TVET) play an important role in the sustainable social development by bridging education and occupational capacity, productivity and decent work. Research on vocational provision and its impact on employment is no longer about the question of to what extent the provision is related to employment, but how the skills system is affected by technological change, local development, workplace learning, and institutional coordination [1–5].

These wider aspects are closely linked to SDG target 4.3 (equal access to technical, vocational, and tertiary education); SDG target 4.4 (relevant technical and vocational skills for employment, decent work, and entrepreneurship); SDG target 4.a (effective and inclusive learning environments); and SDG target 4.c (qualified teachers). In the context of this policy discourse, the Chinese vocational system has

grown, but research efforts have been ongoing to uncover issues of status stratification, non-strong signaling of qualifications, lack of high-level skills, and conflict between expansion and quality [6–9]. Training bases are at the heart of these concerns as they are the physical, digital and organizational environment in which practical learning, involvement in enterprises and competence building is expected. It is important to note that the quality of their work can't just be assumed from their number in the facilities: A "well-equipped base" could be underutilized, the quality of enterprise involvement could be inconsistent, and higher participation could lead to better transition success.

There are four key questions that are addressed by the Context-Input-Process-Product (CIPP) model: What are the needs and environmental conditions? (Context) Inputs: What resources are available? How activities are put into practice (Process)? What are the results of the operation (Product) [10,11]? The four dimensions in this study are not considered to be a set sequence of events nor to be independent of one another as a single score. Rather they are employed to create evidence and provide feedback for planning. Stakeholder involvement, teacher / trainer competence and career guidance have been highlighted as key factors in vocational education in the previous studies [12–14]. Other elements of effective vocational provision that have been identified are the quality of the curriculum, the involvement of industry and the evaluation's transparency [15–18]. Also, research in the area of digital repository and smart learning environment indicates that technology becomes useful when it is directly linked to the pedagogical objectives and the real learning activities [19–21]. Research exploring vocational learning, employability and learner development also indicates that the results of educational attainment rely on more than just technological and institutional resources [22–24]. The assessment practices and the arising of related vocational competences are also still relevant in the evaluation of vocational provision's effectiveness [25]. Results from the various applications of immersive and virtual-reality technologies indicate that simulation technologies can be used to enhance vocational learning, however, the effectiveness of the technology relies on proper instructional design and the engagement of the learner, as well as the meaningful assessment [26–28]. Studies on mixed reality synthesis in recent years also suggest that the benefits of mixed reality approaches include cognitive, behavioral, and affective benefits; however, the results are inconsistent when considering the various implementations [29]. Meanwhile, research into the situation and appeal of vocational education warns that growth in the system does not automatically bring about an enhanced position in society or greater perceived value [30,31]. A study on the quality of the internship shows that what a learner actually experiences in the workplace is important as well as how many internships he or she has completed [32]. Other research with youth at risk of social exclusion also reveals that greater access does not always lead to equal participation or learning outcomes [33]. They are similar to the concerns raised in CIPP evaluations, which focus on the importance of looking at context, resources, implementation, and outcomes together instead of relying on individual indicators [34]. Even with these progressions, there are three related gaps that exist. The majority of CIPP applications are concerned with bounded educational programs, as opposed to providers of a provincial vocational system reporting infrastructure, activities and outcomes across a number of units [10,11,34]. In recent

TVET studies, the integration of components of vocational education into an administrative evaluation system that addresses SDG questions is comparatively less studied, and rather the attention is paid to individual components of TVET. Furthermore, there is little methodological support for how to deal with the overlap and conflict between values from the different sections of the narrative, tables, and standardized appendices so that they do not result in an unsupported composite ranking or a similar causal model.

This study aims to overcome these shortcomings by carrying out a descriptive evaluative case study at the provincial level: Guangdong Provincial Higher Vocational Education Quality Report 2023 [35], and using selected SDG 4 reporting priorities as a lens for interpretation [36]. This focus on training bases is maintained with the 40-indicator core set limited to system conditions, resources, delivery processes and outcomes that directly relate to practical-training provision, such as equipment, workstations, virtual-simulation bases, industry mentors, co-built training centers, training bases that integrate the curriculum certificate, training bases that support internships, training bases that certify, graduates' transitions, and the outcomes of stakeholders. Other provincial indicators were used only where they delineated the operating context or resource envelope of these bases. The study's original contribution is an auditable process that differentiates reported fields from independent observations, maps only available indicators to CIPP and SDG 4, maintains definitional conflicts and separates capacity expansion from reported results. The study addressed three research questions: RQ1: Can provincial quality-reporting data support a transparent CIPP-SDG 4 evaluation of training-base-related provision? RQ2: What strengths and gaps are evident across the Context, Input, Process, and Product dimensions? RQ3: How did comparable indicators change from 2022 to 2023, and do the descriptive findings remain stable after deduplication and source reconciliation?

2. Materials and methods

2.1. Study design, unit of analysis and data source

This study used a descriptive evaluative case-study design based on documentary analysis and secondary analysis of aggregated administrative data. The unit of analysis was Guangdong's provincial higher-vocational education system, not an individual institution, learner, or training center. The 2023 profile and the 2022–2023 trend series were interpreted specifically for training-base-related provision through indicators of practical capacity, finance, staffing, industry participation, curriculum delivery, certification, graduate transitions, and stakeholder outcomes. The source report, published in January 2024, consolidates institutional quality reports and provincial administrative statistics [35]. The analytical record retained the original Chinese indicator labels, English translations, units, source pages, CIPP classifications, SDG 4 links, and definitional conflicts. The report also provided survey-response counts for 527,021 current students, 278,804 graduates, 60,232 staff, 31,640 employers, and 641,613 parents. These counts describe survey coverage and were not treated as independent province-level observations.

2.2. Dataset validation and analytical feasibility

Validation preceded interpretation. The appendix reported 123 indicators, of which 121 contained values, yielding 98.37 % field completeness. The two unreported fields concerned full-time aesthetic-education teachers and class advisers. The indicators included counts, percentages, monetary values, ratios, teaching hours, and per-student measures. Forty indicators were retained in the training-base core set, and none had missing values. The full appendix contained 5 Context, 46 Input, 22 Process, and 50 Product indicators; the core set contained 2, 19, 10, and 9 indicators, respectively. The imbalance reflects the content of the source report rather than a sampling design.

2.2.1. Indicator selection and classification

Five pre-specified criteria were used to designate the core set. Indicators were not included if they referred to activities beyond the scope of the training-base focus, were reported in a numerical value and unit but not a standard appendix table, or had a reported value that was not interpretable, or duplicated a retained measure, or had a reported value that was not interpretable. The change in the number of reported indicators from 123 to 121 completed fields and the introduction of the 40-indicator core set, the 37-row comparison series and the smaller primary and sensitivity trend sets is described in **Supplementary Tables S1–S4**.

Author team members used the operational definitions given in **Table 1** to select and map the CIPP-SDG 4. There was no external expert panel or Delphi procedure, content-validity ratio, content-validity index, or formal inter-rater agreement test. The manuscript therefore reports rule-based content relevance and traceability, but does not claim formal or psychometric content validity. Ambiguous or conflicting values were not resolved by preferring the larger value, but were kept in the source-discrepancy log and the standardized value in the appendix was used in the primary 2023 profile.

Table 1. Variable definitions and CIPP-SDG 4 measurement mapping.

CIPP dimension	Operational domain	Representative indicators and units	Preferred direction	SDG 4 link
Context	System demand and scale	Full-time enrolment (persons); professional clusters (count); regional coordination mechanisms (count)	Descriptive / higher when operational	4.3, 4.4
Context	Industry alignment	Professional or integration structures linked to industry demand (count)	Higher, subject to active use	4.4
Input	Human resources	Student-teacher ratio; dual-qualified teachers (%); senior-title teachers (%); full-time teachers (persons)	Lower ratio; higher qualifications/capacity	4.c
Input	Finance and protection	Fiscal appropriation per student (CNY); internship funds and insurance (CNY)	Higher, with adequacy benchmark	4.a, 4.4
Input	Physical and digital capacity	Equipment per student (CNY); workstations per student; repositories, courses and simulation bases (count)	Higher, conditional on utilization	4.a, 4.4
Process	Industry participation	Partner enterprises, co-built centers and industry mentors (count); mentor hours	Higher, conditional on quality	4.4, 4.c
Process	Curriculum and delivery	Joint courses/textbooks; curriculum-certificate courses/hours; non-degree training projects/hours	Higher, conditional on learning quality	4.3, 4.4

Table 1. (*Continued*).

CIPP dimension	Operational domain	Representative indicators and units	Preferred direction	SDG 4 link
Product	Skills and transitions	Certificate recipients; destination rate (%); income (CNY/month); three-year promotion (%)	Higher	4.4
Product	Stakeholder outcomes	Student, graduate, employer and parent satisfaction (%)	Higher	4.4, 4.a

Note: Preferred direction records the evaluative interpretation used in the change analysis; it is not a causal sign. The mapping describes training-base-related provision at the provincial-system level and does not imply that each indicator is measured at an individual base. Source: Authors' rule-based mapping from the official report [35].

2.3. CIPP-SDG 4 measurement framework

Context indicators represented system scale, policy conditions, and regional industry demand. Input indicators covered finance, staffing, equipment, digital resources, and practical-training capacity. Process indicators covered enterprise engagement, curriculum-certificate alignment, industry-mentored training, internship support, and non-degree training delivery. Product indicators covered skills certification, graduate transitions, income, career progression, and stakeholder satisfaction. The operational domains are defined in **Table 1**. Indicators were retained in their original units and were not combined into a common total. **Figure 1** shows the sequence of evaluation questions and the feedback function of CIPP; the arrows are not causal pathways.

Figure 1 organizes the analysis as four sequential evaluation questions. Context identifies system needs and conditions, Input identifies resources, Process examines implementation, and Product summarizes reported outcomes. The feedback arrow represents the use of findings for subsequent planning rather than an estimated causal effect.

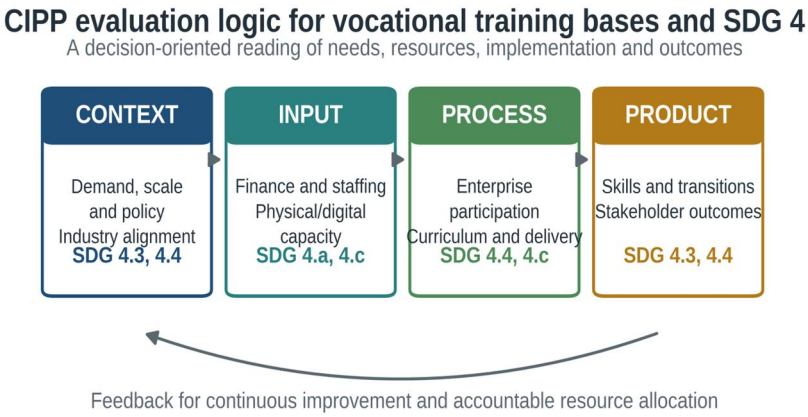

Figure 1. CIPP-based evaluation logic and SDG 4 linkages used in this study. Arrows show the sequence of evaluation questions and feedback, not estimated causal effects. Source: Authors' framework based on [35,36].

2.4. Change analysis and sensitivity tests

For each comparable indicator *i*, absolute change was calculated as

$$\Delta_i = X_i, 2023 - X_i, 2022 \quad (1)$$

Relative change was calculated as

$$r_i = (X_i, 2023 - X_i, 2022)/|X_i, 2022| \quad (2)$$

Direction-adjusted change was defined as

$$d_i = s_i r_i, \qquad s_i \in \{+1, -1\} \quad (3)$$

where $s_i = +1$ when a higher value was preferred and $s_i = -1$ when a lower value was preferred. In plain terms, direction adjustment expresses improvement consistently: Increases receive a positive sign when higher values are preferred, while decreases receive a positive sign when lower values are preferred, as for the student-teacher ratio. Indicators with a zero baseline were described as new structures and were not assigned a relative rate. Cohort-size counts were treated as descriptive when an increase did not necessarily represent better quality. Dimension-level medians, means, and positive / negative counts were used only as descriptive summaries of the available indicators. They are not standardized effectiveness scores or directly comparable estimates because the dimensions contain different numbers, units, and types of indicators.

The comparison file contained 37 reported rows. The primary analysis excluded four duplicate or definitionally overlapping rows from a secondary industry-academia table. It also treated three zero-baseline structures as new developments and one graduate-cohort count as descriptive rather than direction-scored. The resulting primary set contained 29 direction-scored indicators: 15 Input, 5 Process, and 9 Product indicators. The sensitivity analysis reintroduced the four overlapping rows, producing 33 direction-scored indicators: 16 Input, 8 Process, and 9 Product indicators. A separate source-reconciliation check preserved alternative values from narrative tables and appendices rather than forcing a single harmonized estimate. Examples included 447 versus 450 virtual-simulation bases, 58,404 versus 58,998 full-time teachers, and 216 enterprise co-built practice centers versus 258 total open regional practice centers. The standardized appendix value was used for the primary 2023 profile because it is the report's common reporting table; alternative values remained documented.

2.5. Reproducibility, ethics and scope control

The overall data-processing pipeline is summarized in **Figure 2**. The process starts with source extraction and validation, followed by application of rules for indicator selection, and continues with CIPP-SDG 4 mapping and indicator-level trend analysis and source-discrepancy checks. The analysis package contains the entire appendix, the 40-indicator core set, the comparison series, the calculation formulas, a data dictionary and a source-discrepancy log. The indicator-selection matrix is listed in Supplementary **Tables S1–S3** and all comparison rows are reported in Supplementary **Table S4.** There was no intervention, identifiable records or direct contact with participants, and no institutional ethical approval or informed consent. The analysis is limited to Guangdong and is not a causal estimate of the effect of any

specific training base but a 'pure' system-level assessment with training-base implications.

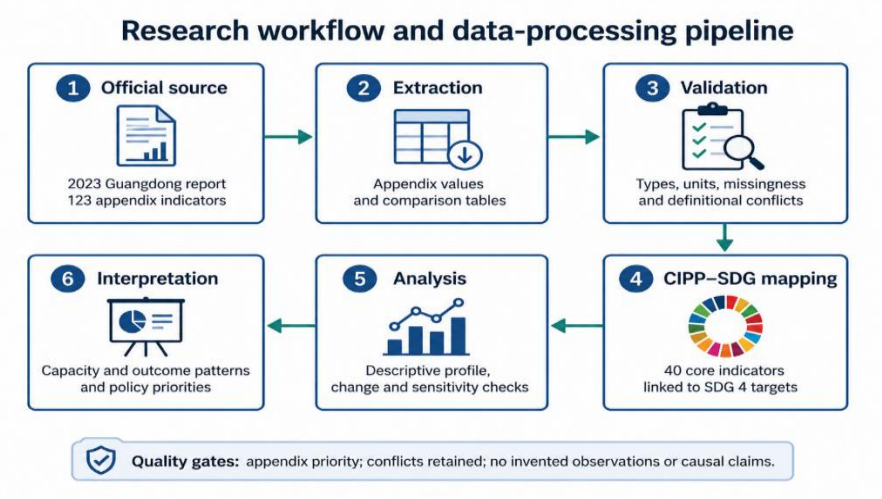

Figure 2. Research workflow and data-processing pipeline. The quality gates preserve source conflicts and prevent invented observations or causal claims.

3. Results and discussion

3.1. Dataset adequacy and measurement validity

The validated descriptive profile is shown in **Table 2**. The report was fully completed with sufficient information for a structured CIPP case evaluation; 121 of the 123 fields in the appendix were filled out; and all 40 core indicators were reported. The indicators stayed in their original units since percentages, monetary values, counts, ratios and teaching hours are not commensurable. It would be misleading to combine them into a total, as this would mask substantive differences and give false precision. Therefore, indicator-level profiles and direction-adjusted change within similar series were analyzed.

Table 2. Dataset validation and selected 2023 descriptive indicators.

Validation or indicator	2023 value	Unit / coverage	Interpretive note
Appendix completeness	121/123 (98.37 %)	reported fields	Two values were not reported.
Core CIPP dataset	40	indicators	No missing values; province-level profile.
Observation level	1	province-year	Not 40 independent cases.
Current-student survey	527,021	respondents	Source survey coverage.
Graduate survey	278,804	respondents	Source survey coverage.
Employer survey	31,640	respondents	Source survey coverage.
Full-time enrolment	1,225,517	persons	System scale.
Student-teacher ratio	16.79	students per teacher	Appendix value; alternative table reports 19.23.
Dual-qualified teachers	63.40	%	Appendix value.
Teaching-resource repositories	853	count	Digital / curriculum input.
Virtual-simulation bases	450	count	Narrative value is 447.

Table 2. (*Continued*).

Validation or indicator	2023 value	Unit / coverage	Interpretive note
Enterprise co-built practice centers	216	count	Narrative total of open centers is 258.
Industry mentors	19,359	persons	3,190,770.22 annual teaching hours.
Skill-certificate recipients	109,182	persons	8.91 % of full-time enrolment.
Graduate destination rate	92.16	%	Product outcome.
Employer satisfaction	98.16	%	Appendix value; figure reports 98.18 %.

Note: Standard deviations are not reported because the unit of analysis is one provincial system. Survey response counts are coverage measures, not independent province-level observations. Source: [35].

Table 3 separates the concepts of rule-based content relevance from psychometric validity. Explicit mapping of each source definition to the operational CIPP domains and the SDG 4 targets increased the relevance, and the original label, unit, source page and discrepancy notes also enhanced the traceability of the sources. The mapping is not a formal expert validation or psychometric content validation. Because the administrative indicators are not repeated items measuring a common latent construct, it was not possible to construct reliability or convergent or discriminant validity. It would therefore be misleading to report a correlation matrix or an alpha coefficient.

Table 3. Measurement adequacy, traceability, and validity limits.

Assessment	Result	Scientific interpretation
Missing-value analysis	2 of 123 appendix indicators missing; 0 of 40 core indicators missing	Adequate for descriptive evaluation; no imputation was applied.
Content relevance	Supported by rule-based mapping	Indicators were mapped by explicit source definitions; no formal expert-validation or inter-rater procedure was conducted.
Traceability	Supported	Original label, unit, report page and discrepancy notes were retained.
Correlation analysis	Not estimable	One provincial observation does not permit a meaningful correlation matrix.
Internal consistency / reliability	Not applicable	Administrative indicators are not interchangeable items of a latent scale.
Convergent and discriminant validity	Not estimable	No repeated construct items or institution-level sample was available.
Criterion validity	Partially supported	Outcomes are official indicators, but independent external verification was unavailable.

Note: “Not estimable” is a design-based conclusion. No synthetic correlation, reliability or validity statistics were generated.

3.2. Context and input profile

The context evidence, in turn, portrays a large, differentiated system: 1,225,517 full-time students, 604 professional clusters, and a considerable amount of policy activity regarding industry / education integration. The report for 2023 noted that there were 28 applications for the municipal industry-education consortia, seven applications for sectoral integration communities and 103 applications submitted for the open regional practice centers [35]. The figures reflect the quick institutionalization of coordination mechanisms; applications are not necessarily the same as quality of operation or successful utilization.

There was a great amount of input capacity. The student-teacher ratio in the standardized appendix was 16.79, 63.40 % dual-qualified full-time teachers, and 25.27 % full-time teachers with senior professional titles, with a total number of

58,998 full-time teachers. The digital resources consisted of 853 teaching-resource repositories, 7196 premium online courses, 450 virtual-simulation bases, 0.77 teaching and research equipment workstations on campus per student, and CNY 12,985.24 in teaching and research equipment per student. The total value of the practical equipment provided by enterprises is CNY 934.07 million, based on the reported CNY 93,407.28 ten-thousand-unit value. **Figure 3a** shows some of the physical and digital resource counts and **Figure 3b** shows stakeholder-outcome and teacher-quality percentages. The two panels are not on a common scale nor are they a composite index.

The overall trend of the input profile was not consistently favorable. The annual special fiscal appropriation grew by CNY 1.488 billion to CNY 8.780 billion, while the fiscal appropriation per student dropped by CNY 2663.65 to CNY 17,471.12, a decrease of 13.23 %. This is a difference, meaning that aggregate budget growth was not in line with reported scale of the system. In contrast, the per-student fiscal pressure did not seem to affect internship-specific funding, which increased by 26.87 %, or internship liability insurance funding, which increased by 29.94 %, indicating specific protection of work-based learning under per-student fiscal pressure.

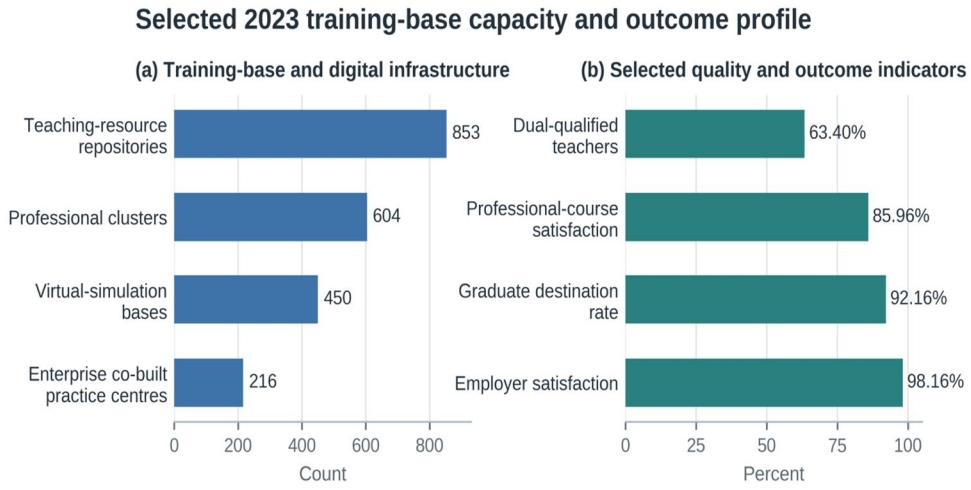

Figure 3. Selected 2023 training-base capacity and outcome profile: **(a)** infrastructure counts; **(b)** selected teacher-quality and outcome percentages. Source: Authors' calculations from [35].

3.3. Process performance

The primary trend set had the most consistent description improvement as process indicators. Enterprise partners increased from 24,375 to 31,005 (27.20 %), cooperative majors from 2661 to 3030 (13.87 %), jointly developed courses from 9630 to 11,410 (18.48 %), and jointly compiled textbooks from 1028 to 1760 (71.21 %). The biggest relative change was in jointly prepared textbooks, but the number of these is not a measure of instructional quality. In 2023, there were 18614 integrated curriculum certificate courses and 1075722 teaching hours of integrating curriculum and certificate courses.

Collaborations were not limited to formal ones. Approximately 164.82 hours per mentor were spent by industry mentors (19,359) delivering a total of 3,190,770.22

hours of teaching. Industry mentors (19,359), comprising 65 master craftsmen or model workers, delivered 3,190,770.22 hours of teaching, which is equivalent to 164.82 hours of teaching per mentor. The total amount of mentors' remuneration was CNY 220.89 million, and 216 open regional industry-education practice centers were reported as jointly built by enterprises. The report also indicated 7329 non-degree training projects and 1,954,796.10 training hours with an average of 266.72 hours per project. These measures do not indicate either attendance quality, equipment use, observed instruction quality or learning gains for learners.

3.4. Product outcomes

Product indicators documented high 2023 levels but mixed trends. Overall, 109,182 learners received vocational skill-level or vocational qualification certificates, equivalent to 8.91 % of full-time enrolment reported in the appendix. The graduate destination rate was 92.16 %, average monthly income was CNY 4190, and the three-year promotion rate was 48.95 %. Of 444,730 graduates in initial employment, 199,671 were employed locally, representing 44.90 %. Stakeholder ratings were also high: Current-student satisfaction was 80.55 %, professional-course satisfaction was 85.96 %, graduate satisfaction was 88.32 %, and employer satisfaction was 98.16 %.

High outcome levels did not prevent deterioration in selected trends. The destination rate fell from 98.62 % in 2022 to 92.16 % in 2023, an absolute decline of 6.46 percentage points and a relative decline of 6.55 %. The source table printed a change of -3.20, which is arithmetically inconsistent; that figure was retained in the discrepancy log but not used analytically. Three-year graduate satisfaction and employer satisfaction increased slightly, parent satisfaction fell by 0.69 percentage points, and monthly income rose by 6.89 %. The reported graduate cohort nearly doubled, increasing by 91.46 %. The pattern is therefore consistent with slower movement in several outcomes than in selected capacity and process indicators, but it may also reflect cohort expansion, labor-market conditions, reporting changes, indicator sensitivity, or time lags between investment and outcomes.

3.5. Main directional results

The primary deduplicated results are summarized in **Table 4** and **Figure 4**. Process indicators had the largest median direction-adjusted change (18.48 %), and all five moved in the preferred direction. Among the 15 Input indicators, 14 moved in the preferred direction and one moved in the opposite direction; the median and mean changes were 10.07 % and 10.71 %, respectively. Product indicators included seven positive and two negative series, with a median of 1.52 % and a mean of 11.07 %. The higher mean was influenced by the large increase in confirmed graduate destinations as the graduate cohort expanded. For this reason, the median provides the more stable description of typical Product movement.

Context trends were not summarized as relative rates because two Context coordination mechanisms and one Process-related practice-center application measure had zero baselines in 2022. They were reported as new institutional structures rather than assigned mathematically undefined growth rates. This treatment preserves the distinction between new governance arrangements and measured outcome

improvement.

Table 4. Main direction-adjusted 2022–2023 evaluative results.

CIPP dimension	Primary indicators (n)	Median direction-adjusted change	Mean direction-adjusted change	Positive / negative	Interpretation
Context	2 new structures	Not calculated	Not calculated	2 new / 0 closed	Zero-baseline Context structures were described, not converted to rates.
Input	15	10.07 %	10.71 %	14 / 1	Broad capacity gains; per-student fiscal support declined.
Process	5 scored + 1 new	18.48 %	27.18 %	5 / 0; 1 new	All scored indicators improved; one zero-baseline structure was reported separately.
Product	9	1.52 %	11.07 %	7 / 2	Typical gains were modest; mean affected by cohort-scale growth.

Note: Positive / negative counts refer to the preferred evaluative direction among scored indicators. Two Context structures and one Process structure had zero baselines and were reported as new developments rather than assigned relative rates.

Figure 4 displays the dimension medians and means for the direction-scored primary set. The bars are descriptive summaries; they are not confidence intervals, significance tests, or causal effect estimates.

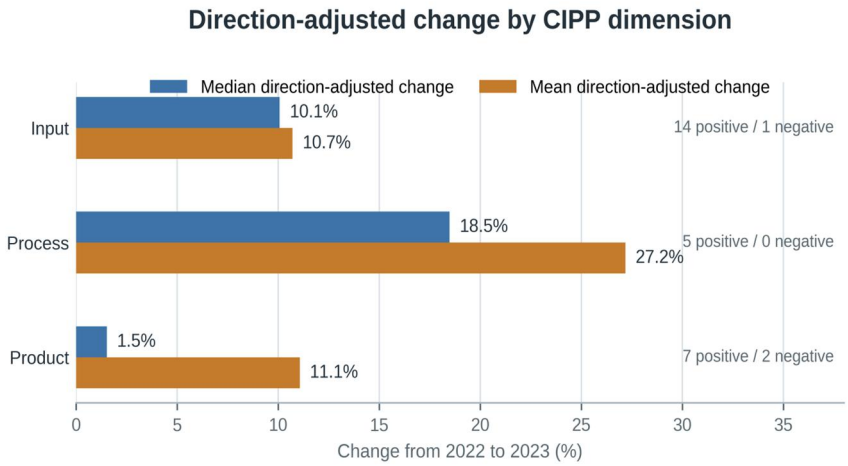

Figure 4. Direction-adjusted change in the primary deduplicated indicator set, 2022–2023. Bars show dimension medians and mean; labels report positive and negative series.

3.6. Robustness and sensitivity

When all comparison rows were added back in, the descriptive ordering stayed the same. **Table 5** indicates that the Process median was still positive at 16.18 %, and all 8 had a Process series in the desired direction. The Input median was reduced from 10.07 % to 8.32 % due to overlapping faculty statistics being included in the extended set from a different reporting table, and 14 of 16 Input series stayed positive. The outcomes of the products remained the same. These checks do not provide statistical significance or causation, rather they contribute to the stability of the descriptive pattern that resource and partnership indicators grew faster than the traditional Product indicator.

The descriptive interpretation was not changed, although some point estimates

changed due to source reconciliation. The numbers of virtual-simulation bases reported in the source-report appendix were 450, while the numbers in the narrative were 447; the numbers of full-time teachers in the appendix were 58,998, while in the narrative they were 58,404; the percentages of employer satisfaction in the appendix were 98.16 %, while in the narrative they were 98.18 %. These differences are statistically insignificant but in terms of repeatability they are significant. The 216 enterprise co-built practice centers and the 258 total open regional centers are not the same thing, but rather two different definitions. The labels were therefore kept apart and no conflicts in the values were averaged.

Table 5. Robustness, sensitivity and source reconciliation analysis.

Sensitivity specification	Input result	Process result	Product result	Conclusion
Primary deduplicated set	*n* = 15; median 10.07 %; mean 10.71 %; 14+/1-	*n* = 5; median 18.48 %; mean 27.18 %; 5+/0-	*n* = 9; median 1.52 %; mean 11.07 %; 7+/2-	Process indicators showed larger typical descriptive changes than Product indicators.
All comparison rows	*n* = 16; median 8.32 %; mean 9.62 %; 14+/2-	*n* = 8; median 16.18 %; mean 22.24 %; 8+/0-	Unchanged	The descriptive ordering remained stable.
Appendix-priority values	450 virtual bases; 58,998 teachers; 98.16 % employer satisfaction	216 enterprise co-built centers	92.16 % destination rate	Standardized appendix used for 2023 profile.
Narrative / alternative values	447 virtual bases; 58,404 teachers; 98.18 % employer satisfaction	258 total open centers	Destination-rate source change label inconsistent	Alternative values were retained; the descriptive interpretation did not depend on small differences.
Scale-only counts excluded	Cohort size treated as descriptive	New zero-baseline structures not rate-scored	Confirmed-destination growth interpreted with cohort context	This treatment reduced the risk of overstating quality improvement.

Note: "+" and "-" indicate positive and negative direction-adjusted series. Alternative values were retained in the discrepancy log rather than averaged.

4. Discussion

4.1. Interpreting capacity, processes, and outcomes

The primary outcome is a descriptive trend that is quite consistent, where most of the selected Input and Process indicators rose more quickly than the typical Product indicator. This pattern is consistent with there being a mismatch between the increase in selected capacities and measures of activity and a movement of some outcomes, though not that the resources were used ineffectively, or that outcomes worsened due to expanded activity. The possibility of cohort expansion, labor-market conditions, reporting changes, indicator sensitivity and time lags between investment and graduate outcomes are still possible explanations. The findings contribute to the application of CIPP in TVET, as the four dimensions can be followed as an evidence-organizing process: Context is used to identify needs and conditions, Input is used to document mobilized resources, Process is used to document implementation, and Product is used to document reported outcomes. This sequence is not a linear causal model, nor are the four dimensions comparable components of a single effectiveness score.

This is in line with studies that seek to differentiate between workplace exposure and workplace curriculum quality [5] and that highlight trainers, guidance and learner experience [12–15]. It is also consistent with efforts to involve industry in the education process [16,17]. The number of partnerships, courses, or textbooks can

increase opportunities and their learning value relies on co-design, instructional responsibility, assessment, use, and experience. The Guangdong report thus provides a clearer picture of the architecture and scale of collaboration than its pedagogical quality.

4.2. Methodological contribution

The methodological contribution is a defensible method for analyzing administrative reports with very few statistical observations, but with many dimensions. The report contains 123 fields, but these fields do not constitute 123 independent cases or observations. This way, pseudo-replication from correlations of unrelated aggregates, machine-learning models built on a single province-year, or reliability statistics for administrative variables not intended to be a scale is avoided. The content relevance and traceability that can be added with rule-based mapping and the discrepancy log increases auditability, but can't replace formal expert validation.

The procedure builds on the systematic CIPP reviews and program evaluations [10,11,34], and is an extension of the same issue identified in public-sector datasets: A single construct can be presented in multiple ways, as narrative text, in a thematic table, or in a standardized appendix with slightly different values. The inclusion of the standardized appendix while allowing for alternative values is clearer than choosing the more favorable of several different definitions or the average value of such incompatible values. The direction-adjusted measure also clarifies the direction without suggesting causality. Because the dimensions are composed of diverse indicators, the medians continue to describe the dimensions rather than provide estimates of comparative effectiveness.

4.3. SDG 4, policy, and managerial implications

The results have specific implications for the chosen SDG 4 targets. Access-related system and non-degree training indicators are associated with target 4.3; certification, industry mentoring, curriculum-certificate integration, employment outcomes, and other skills-related indicators are associated with target 4.4; equipment, virtual-simulation bases, workstations, and internship protection are associated with target 4.a; and teacher ratios, dual-qualified staff, and industry mentors are associated with target 4.c. Learner-level subgroup data are not available in the source. The study thus assesses selected SDG 4 priorities, instead of SDG 4 in its entirety.

Facility counts should be supplemented with measures of utilization and learning quality for provincial and institutional managers. Each physical and virtual base should be reported separately for annual reporting of learner contact hours, occupancy/utilization rate, downtime, completion, assessment performance, instructor qualifications, enterprise co-teaching intensity and learner feedback. Task authenticity and learning outcomes should be used to evaluate digital and immersive facilities, not strictly on the basis of the number of facilities installed [19,26–29].

The adequacy of resources for individual students must also be monitored. While the amount of money appropriated per student fell, the amount of money appropriated for aggregates grew. It is important that the expansion goals go along with minimum per learner funding adequacy thresholds, equipment renewal and maintenance plans,

mentor capacity development, and internship insurance coverage. Reported number of practical workstations per student (0.77) and enterprise-provided practical equipment per student (approx. CNY 762) are useful benchmarks, but it is important to have program-level data on need and utilization to interpret.

Outcomes should be reported in a longitudinal and disaggregated way. The institution, program cluster, sex, rural or urban origin, disadvantage of the student and the employer should all be reported for destination rates, income, promotion and satisfaction of the employer. Because these subgroup variables are unavailable, the present study cannot evaluate SDG target 4.5. This is an important omission because social sorting and exclusion are important factors determining participation and status in the labor market and formal access to it [6,7,30–33]. Reporting on enterprise collaboration should also separate the signs of an agreement from a verified delivery, co-teaching, assessment responsibility and an independently validated result.

4.4. Limitations and future research

There are six limitations in the study. First, it is looking at one province and is not able to be generalized nationally. Second, the main profile is cross-sectional and the comparison window includes only two years, and therefore short-term (temporary) cohort, policy or labor-market conditions cannot be disentangled from longer-term trends. Third, the data are aggregated and do not identify mechanisms at learner, program, institution or individual-level. Fourth, the study is based upon an official administrative report, which could include institutional self-reporting, selective reporting, or reporting incentives. Fifthly, the raw data from which the estimates were derived were not available for independent verification, nor were the denominators of the surveys, and the methods used for measuring population change were unavailable. Sixth, there are differences in values by report locations and / or report definitions, which suggests some report and/or transcription errors. Sampling uncertainty was not quantified in the sensitivity analysis because of duplication and alternative source values. The challenge of SDG target 4.5 is also not being addressed due to the absence of disaggregated subgroup data. These limits do not preclude the use of descriptive evaluation, but they limit causal inference, ranking and generalizing.

Future studies should build up a harmonized institution-year panel from the individual college quality reports. The training base should have consistent identifiers and definitions of utilization hours, learner participation, instructor qualifications, enterprise co-teaching, assessment results, completion, equipment downtime and graduate outcomes. Outcomes need to be tracked for a minimum of three years and broken down into institutional, program, sex, rural/urban, disability status and socioeconomic disadvantage. This classification framework should also be evaluated by independent vocational-education and evaluation experts using a documented consensus or content-validation procedure. Multi-level models taking into consideration the size of the institution, the program mix, the governance status and regional economic conditions could benefit from such a panel. Machine learning would be suitable once there was an institution-year sample size large enough and consistently defined, and train-test separation and explainability in place. The numerical results cannot be extrapolated to other provinces or vocational systems that

provide similar quality reports, but the process of source validation and indicator mapping could be adapted for other provinces or vocational systems.

5. Conclusion

The research findings indicate that there are high levels of reported capacity for practical and industry-related education in the Higher Vocational Education system at the provincial level in Guangdong. From 2022 to 2023, all five direction-scored Process indicators in the primary set moved in the preferred direction, with a median direction-adjusted change of +18.48 %; the typical Product movement was smaller, while the graduate destination rate and fiscal appropriation per student fell. A pattern of faster growth in selected capacities and process measures and slower growth in some outcomes does not indicate that resources contributed, or did not contribute, to specific outcomes. The analysis covers selected indicators related to SDG targets 4.3, 4.4, 4.a, and 4.c, but not target 4.5 because the source lacks disaggregated subgroup data. The study contributes to the field of educational evaluation by demonstrating the utility of CIPP in structuring a diverse body of administrative data without assuming the dimensions of the model of evaluation are a causal sequence or an effectiveness continuum. Implementation should include the assignment of persistent identifiers to training bases, consistent definitions of what is being reported, tracking of annual utilization and tracking of instructional quality, and the creation of linkages with disaggregated outcomes for graduates of training bases extending over a minimum of three years. The numerical findings should not be statistically extrapolated beyond Guangdong; however, the auditable mapping and source-reconciliation process may be adapted to similar provincial reports. Future studies should validate the classification with independent experts and then create a harmonized panel at the institutional level, prior to applying multilevel modelling or predictive approaches.

Author contributions: Conceptualization, XS and PC; methodology, XS; software, PC; validation, XS, XS and PC; formal analysis, XS; investigation, XS; resources, XS; data curation, XS; writing—original draft preparation, XS; writing—review and editing, XS; visualization, XS; supervision, XS; project administration, PC; funding acquisition, PC. Both authors have read and agreed to the published version of the manuscript.

Funding: None.

Ethical approval: Not applicable.

Informed consent statement: Not applicable.

Data availability statement: The source report is publicly available at https://www.tech.net.cn/upfiles/zlbg2023_new/sf/gds.pdf. All values used in the analysis are traceable to the cited report, and the calculation logic is reported in the manuscript.

Supplementary materials: Tables S1–S3 provide the full 123-indicator selection and classification matrix, including original labels, English translations, units, source pages, CIPP dimensions, SDG 4 links, preferred directions, and reasons for inclusion

or exclusion. **Table S4** reports all 2022–2023 comparison rows, calculation fields, primary-set status, sensitivity-set status, and source-discrepancy notes.

Acknowledgements: The authors thank the Guangdong Provincial Department of Education for making the provincial higher vocational education quality report publicly available.

Conflict of interest: The authors declare no conflict of interest.

References

1. McGrath S, Yamada S. Skills for development and vocational education and training: Current and emergent trends. International Journal of Educational Development. 2023; 102: 102853. doi: 10.1016/j.ijedudev.2023.102853
2. McGrath S, Ramsarup P, Zeelen J, et al. Vocational education and training for African development: A literature review. Journal of Vocational Education & Training. 2020; 72(4): 465–487. doi: 10.1080/13636820.2019.1679969
3. Li J, Pilz M. International transfer of vocational education and training: A literature review. Journal of Vocational Education & Training. 2023; 75(2): 185–218. doi: 10.1080/13636820.2020.1847566
4. Rözer J, van de Werfhorst HG. Three worlds of vocational education: specialized and general craftsmanship in France, Germany, and the Netherlands. European Sociological Review. 2020; 36(5): 780–797. doi: 10.1093/esr/jcaa025
5. Rintala H, Nokelainen P. Vocational education and learners' experienced workplace curriculum. Vocations and Learning. 2020; 13(1): 113–130. doi: 10.1007/s12186-019-09229-w
6. Wang G. 'They are bad seeds': stereotyping habitus in Chinese vocational education and training colleges. British Journal of Sociology of Education. 2021; 42(7): 1008–1021. doi: 10.1080/01425692.2021.1959299
7. Guan S, Blair E. Exploring conceptualisations of vocational education in China: hierarchical education system mirrors social hierarchy. British Journal of Sociology of Education. 2024; 45(5): 778–797. doi: 10.1080/01425692.2024.2361283
8. Song X, Xu D. More graduates, fewer skills? Vocational education expansion and skilled labour shortages in China. The China Quarterly. 2024; 260: 970–985. doi: 10.1017/S0305741023001856
9. Xue E, Li J. Exploring macro education policy design on vocational education system for new generation migrant workers in China. Educational Philosophy and Theory. 2020; 52(10): 1028–1039. doi: 10.1080/00131857.2019.1675468
10. Toosi M, Modarres M, Amini M, Geranmayeh M. Context, input, process, and product evaluation model in medical education: A systematic review. Journal of Education and Health Promotion. 2021; 10: 199. doi: 10.4103/jehp.jehp_1115_20
11. Ratnaya G, Indriaswuri R, Widayanthi DGC, et al. CIPP evaluation model for vocational education: A critical review. Education Quarterly Reviews. 2022; 5(3): 1–8. doi: 10.31014/aior.1993.05.03.519
12. Boonk LM, Ritzen HJM, Gijselaers HJM, Brand-Gruwel S. Stimulating parental involvement in vocational education and training (VET): A case study based on learning histories of teachers, principals, students, and their parents. Teaching and Teacher Education. 2021; 100: 103279. doi: 10.1016/j.tate.2021.103279
13. Schmid E, Garrels V. Parental involvement and educational success among vulnerable students in vocational education and training. Educational Research. 2021; 63(4): 456–473. doi: 10.1080/00131881.2021.1988672
14. Magee M, Kuijpers M, Runhaar P. How vocational education teachers and managers make sense of career guidance. British Journal of Guidance & Counselling. 2022; 50(2): 273–289. doi: 10.1080/03069885.2021.1948970
15. Wuttke E, Heinrichs K, Koegler K, Just A. How training quality, trainer competence, and satisfaction with training affect vocational identification of apprentices. Frontiers in Psychology. 2024; 15: 1200279. doi: 10.3389/fpsyg.2024.1200279
16. Rohr-Mentele S, Holtsch D. Recording business transactions in textbook and receipt form. Vocations and Learning. 2022; 15(2): 209–239. doi: 10.1007/s12186-022-09285-9
17. Zhang YB, Wang JB, Shen XD, Song JY. Research on the influence mechanism of enterprises' participation in school–enterprise cooperation based on the analysis framework of the theory of planned behavior. Frontiers in Psychology. 2022; 13: 860045. doi: 10.3389/fpsyg.2022.860045
18. Liu X, Wang K, Luo D, et al. Studies on the quantification and time-sequence development of integration of industry and education level in higher vocational education—based on the panel data of China's provincial from 2016 to 2022. Journal of Infrastructure Policy and Development. 2024; 8(6): 6292. doi: 10.24294/jipd.v8i6.6292

19. Yang C, Kaiser F, Tang H, et al. Sustaining the quality development of German vocational education and training in the age of digitalization: challenges and strategies. Sustainability. 2023; 15(4): 3845. doi: 10.3390/su15043845
20. Kantova R, Motycka V. Evaluation of sustainability of Czech vocational education and practical training for the position of construction manager. Sustainability. 2023; 15(19): 14631. doi: 10.3390/su151914631
21. Hurskainen J, Wenstrom S, Uusiautti S. A PRIDE-theory-based analysis of a positive learning environment in a Finnish vocational education and training institution. Nordic Journal of Vocational Education and Training. 2023; 13(2): 74–99. doi: 10.3384/njvet.2242-458X.2313274
22. Ye JH, He Z, Bai B, Wu YF. Sustainability of technical and vocational education and training (TVET) along with vocational psychology. Behavioral Sciences. 2024; 14(10): 859. doi: 10.3390/bs14100859
23. Wongmonta S. Revisiting wage effects of vocational education and training (VET) over the life cycle: the case of Thailand. International Journal of Educational Development. 2023; 103: 102886. doi: 10.1016/j.ijedudev.2023.102886
24. Zhang YB, Song JY. An empirical study on willingness and behavior of higher vocational students to participate in modern apprenticeship: Based on the theory of planned behavior. SAGE Open. 2024; 14(2): 21582440241252291. doi: 10.1177/21582440241252291
25. Morselli D. To assess or not to assess an entrepreneurship competence in vocational education and training? Results from a case study of Italy. Education + Training. 2024; 66(10): 29–41. doi: 10.1108/ET-06-2023-0242
26. Kablitz D, Conrad M, Schumann S. Immersive VR-based instruction in vocational schools: Effects on domain-specific knowledge and wellbeing of retail trainees. Empirical Research in Vocational Education and Training. 2023; 15: 9. doi: 10.1186/s40461-023-00148-8
27. Chiang S, Chiang D, Chang SH, Yao KC. Innovative virtual reality teaching for sustainable development of vocational high school students: A case study of hair braiding. Sustainability. 2025; 17(9): 3945. doi: 10.3390/su17093945
28. Long Y, Zhang X, Zeng X. Application and effect analysis of virtual reality technology in vocational education practical training. Education and Information Technologies. 2025; 30: 9755–9786. doi: 10.1007/s10639-024-13197-7
29. Bödding R, Schriek SA, Maier GW. A systematic review and meta-analysis of mixed reality in vocational education and training: Examining behavioral, cognitive, and affective training outcomes and possible moderators. Virtual Reality. 2025; 29: 44. doi: 10.1007/s10055-025-01118-z
30. Pilz M. Status, image, standing and so on: A conceptual framework to systematise attractiveness of vocational education and training. Education + Training. 2025; 67(3): 283–298. doi: 10.1108/ET-01-2024-0038
31. Zoellner D. Theorising the status of VET from the institutional logics perspective. Journal of Vocational Education & Training. 2025; 77(4): 939–957. doi: 10.1080/13636820.2024.2369846
32. Yan C. How do vocational students evaluate vocational education internships – an evaluation scale development in China. Education + Training. 2025; 67(1): 90–106. doi: 10.1108/ET-08-2023-0320
33. Bruin MG, Tutlys V, Umarik M, et al. Participation and learning in vocational education and training: A cross-national analysis of the perspectives of youth at risk for social exclusion. Journal of Vocational Education & Training. 2025; 77(3): 706–727. doi: 10.1080/13636820.2023.2283745
34. Sankaran S, Saad N. Evaluating the bachelor of education program based on the context, input, process, and product model. Frontiers in Education. 2022; 7: 924374. doi: 10.3389/feduc.2022.924374
35. Guangdong Provincial Department of Education. Guangdong Provincial Higher Vocational Education Quality Report (2023) (in Chinese). 2024. Available online: https://www.tech.net.cn/upfiles/zlbg2023_new/sf/gds.pdf (accessed on 18 July 2026).
36. Ministry of Education of the People's Republic of China. China SDG 4 Mid-Term Review Report. 2024. Available online: https://articles.unesco.org/sites/default/files/medias/fichiers/2024/04/China %20SDG %204 %20MTR %20Report_FINAL_ENG.pdf (accessed on 18 July 2026).

Appendix. Source-report appendix clarification

In this article, references to the "appendix" in descriptions of source values refer to the standardized appendix of the Guangdong Provincial Higher Vocational Education Quality Report (2023) [35]. Source-report appendix values directly referenced in Section 3.6 are 450 virtual-simulation bases, 58,998 full-time teachers, and 98.16 % employer satisfaction. The full core-indicator listing is provided in Supplementary **Table S2**, and the corresponding source discrepancies are documented in Supplementary **Table S3**.