Implementing Emergency Rescue Skills Training in Urban Rail Transit TVET: A Qualitative Single-Case Study from a Chinese Higher Vocational College

***Liu Siyu[1], Ng Ming Yip[2], Syed Lamsah bin Syed Chear[3]**

[1]Faculty of Education and Social Sciences, Universiti Selangor, Selangor, Malaysia/Chengdu Industry and Trade College, Chengdu, China

[2]Faculty for the Development of Students of Noble Character, Universiti Selangor, Selangor, Malaysia

[3]Faculty of Education and Social Science, Universiti Selangor, Selangor, Malaysia

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.101100064

Received: 09 November 2026; Accepted: 14 November 2026; Published: 20 November 2026

ABSTRACT

China's urban rail transit (URT) network operates in 58 cities, covers 13,067.89 route-kilometres and carries 33.383 billion passenger journeys a year. Because incidents underground are contained or amplified by staff on duty, the capacity to act in the opening minutes is an occupational requirement of the operational workforce rather than a specialist competence. Preparing that workforce is a task of technical and vocational education and training (TVET), yet the emergency component of URT programmes has rarely been examined empirically. This paper reports a qualitative single-case study of how emergency rescue skills training is implemented across the four URT programmes of one Chinese higher vocational college, what the provision does well and what limits it. Data came from interviews with 12 students, four teachers and nine enterprise supervisors, from curriculum documents and from observation of practical sessions, and were analysed thematically. Interpretation was framed by competence-based education, experiential learning theory and the ADDIE instructional-design model. Three challenges emerge. A theory-practice gap concentrates contact hours in classroom delivery and rewards written recall, while scenario drills requiring decision-making, cross-department communication and coordination remain scarce. The training model constrains practice: facilities follow equipment rather than incidents, and several teachers lack recent front-line experience. Industry-education integration is shallow, since enterprises provide placements but seldom co-design modules, co-facilitate drills or co-assess competence. These are not three deficits but one design failure, reproduced by the audit logic of assessment, the visibility of equipment in investment decisions and the contractual scope of order-class partnerships. Institutional commitment, a first-aid certification route and partnerships capable of deeper cooperation are the assets on which reform can build. Colleges should rebuild emergency modules around an explicit competency framework, invest in scenario simulation with structured debriefing, and move enterprise involvement from placement to co-construction. Findings are analytically rather than statistically generalisable.

Keywords: Urban rail transit, emergency rescue skills, TVET, competency-based education, qualitative case study

INTRODUCTION

China operates the largest urban rail transit network in the world. By the end of 2025 the sector served 58 cities across 382 lines and 13,067.89 route-kilometres, carrying 33.383 billion passenger journeys in a single year (China Association of Metros [CAMET], 2026). Those journeys take place largely below ground or on segregated alignment, in stations and trains whose safe operation depends on the coordinated functioning of signalling, traction power, ventilation, platform-edge interfaces and crowd management. Incidents are

infrequent, but their consequences are concentrated: a medical emergency, a stalled train, a fire alarm or a crowd surge is contained or amplified by what staff do in the first few minutes.

The people who act in those minutes are not specialist responders. They are station attendants, train operators, dispatchers and maintenance technicians who happen to be on duty. Research on organisational accidents has established that the decisive variable in such moments is not the completeness of the rule book but the capacity of front-line staff to notice, interpret and contain an emerging deviation before it propagates through a tightly coupled system (Reason, 1997; Weick & Sutcliffe, 2015). Chinese regulation treats the expectation as an obligation rather than a virtue: operators must maintain a three-tier emergency plan system, hold comprehensive and special plan drills at least once every six months, and assess key-post personnel, barring those who fail from duty (Ministry of Transport of the People's Republic of China, 2018). Sector planning places the corresponding training requirement at more than 58 million training activities a year, with full coverage of key posts and at least 150 training hours per person (CAMET, 2023).

Higher vocational colleges supply a large share of this workforce. National vocational education policy requires practical teaching to be prioritised, assessment to be competence-based and employers to participate substantively rather than symbolically (State Council of the People's Republic of China, 2019). The policy is unambiguous about ends; what it cannot specify is the curriculum through which those ends are reached. That curriculum is decided in programme plans, timetables, equipment budgets and assessment regulations, and it is there, rather than in policy text, that the emergency component of a URT programme either acquires substance or thins into a chapter of a textbook.

Whether policy intent reaches that level is largely unexamined, and the shape of the literature explains why. Scholarship on URT safety is dominated by engineering and post-incident analysis, in which the human operator appears as a component of a technical system rather than as a learner. Scholarship on simulation-based safety training is dominated by medicine and aviation, where the instructional evidence is strong but the institutional setting, the resourcing and the professional status of the learner differ sharply from those of a vocational college. Scholarship on TVET in China is dominated by normative argument about what programmes ought to contain, and it is rich in prescription and thin in evidence about what programmes do contain. The consequence is a body of writing in which the policy requirement, the pedagogical ideal and the curriculum document are each well described, and the space between them is not.

This study addresses that space through a single case. It asks three questions: how emergency rescue skills training is currently implemented in the four URT programmes of one Chinese higher vocational college; what strengths and challenges characterise that implementation; and which improvement directions the field evidence supports. The contribution is threefold. Empirically, the study documents a safety-critical curriculum component at the point where policy meets classroom practice, in a sector where the cost of the gap is measured in minutes. Analytically, it shows that the three challenges usually reported separately in this literature are produced by a single design failure and sustained by identifiable institutional mechanisms, a claim that is developed through the discussion rather than asserted in advance. Practically, it identifies leverage points that lie within the authority of a college, which matters because the diagnosis in this field tends to terminate in recommendations that require national reform to implement.

LITERATURE REVIEW

Emergency rescue competence as an occupational requirement

Emergency rescue competence in URT is not a single skill but an integrated capability set. Reviewed against sector practice and training standards, it comprises six dimensions: emergency knowledge; technical response, meaning the correct operation of equipment and procedures under abnormal conditions; situation decision-making under incomplete information; information communication with passengers, colleagues and external agencies; teamwork and coordination across posts and departments; and psychological and adaptive competence, the capacity to act while passengers are distressed and while personal stress is high.

The six dimensions are not separately teachable in any meaningful sense. They are exercised together, under time pressure, in conditions that classroom instruction can describe but cannot reproduce. Reason's (1997) distinction between active failures and latent conditions sharpens the point: what an operator achieves in the first minutes depends less on individual knowledge than on the conditions the organisation has prepared, including procedures that fit the situation, equipment that is where it is needed and colleagues who share a working model of what is happening. Weick and Sutcliffe (2015) add that reliability under uncertainty rests on a set of dispositions rather than a stock of rules, among them a reluctance to simplify, a sensitivity to operations and a commitment to resilience. Competence, on this reading, is relational: it is distributed across people, artefacts and routines rather than stored in a syllabus.

Two implications follow, and they pull against each other. The first is that if competence is integrative and performance-based, its development depends on repeated supervised exposure to realistic conditions, and its assessment on observable performance rather than recall. Where provision is organised around subject content and examined through written tests, the resulting capability is likely to be declarative: students can describe the correct sequence without having practised it. The second is that a six-dimension list, however accurate, is an analytic convenience that invites exactly the fragmentation it is meant to prevent. Once dimensions are written as separate learning outcomes, they can be timetabled, taught and examined separately, and the integration that defines the occupational requirement disappears from the curriculum at the moment the curriculum is documented. This tension between integration as a property of practice and decomposition as a requirement of administration recurs throughout the findings.

From six dimensions to learning outcomes and assessment criteria

The six-dimension model remains an analytic convenience until it is translated into outcomes that can be taught and criteria that can be assessed. Constructive alignment requires that intended outcomes, learning activities and assessment tasks all target the same performance (Biggs & Tang, 2011); where an outcome names integrated performance but the assessment samples recall, the curriculum teaches recall whatever the objectives declare. Applied to the six dimensions, the translation runs as follows. Emergency knowledge becomes the outcome that a student identifies the applicable disposal procedure for a stated incident and explains the reason for each step. Technical response becomes the outcome that the student operates the relevant equipment and follows the relevant procedure correctly under the abnormal conditions of the scenario. Situation decision-making becomes the outcome that the student selects and justifies a defensible course of action when information is incomplete or priorities compete. Information communication becomes the outcome that the student transmits accurate, timely and protocol-compliant information to passengers, colleagues and external agencies. Teamwork and coordination becomes the outcome that the student performs an assigned role within a multi-post response and hands over cleanly to the next actor. Psychological and adaptive competence becomes the outcome that the student continues to act effectively while passengers are distressed and time pressure is high. Each outcome is written as a performance under stated conditions rather than as a topic to be covered.

The assessment criteria must then be specified at the same level of performance. Gulikers et al. (2004) propose five dimensions along which an assessment task approximates professional practice: the task itself, the physical context, the social context, the form of output and the criteria applied. A written test of emergency procedures is weak on the physical and social context and on the form of output, and it applies criteria of correctness that do not capture decision-making, communication or coordination. Criteria aligned to the six outcomes would instead record, for each scenario, whether the procedure was identified, whether the equipment was operated correctly, whether the decision was defensible given the information available, whether the handover was accurate and timely, whether the student performed the assigned role in coordination with others, and whether effective action was maintained under pressure. Because the six dimensions are exercised together in practice, the appropriate unit of assessment is the incident scenario, not the isolated dimension: a single well-designed scenario elicits several dimensions at once, and the criteria record the quality of each. This is the design step that the present provision has not taken, and it explains why the deficits in practice, in assessment and in enterprise involvement appear together rather than separately.

Table 1: The six dimensions of emergency competence translated into learning outcomes and assessment criteria

Dimension	Learning outcome (performance under stated conditions)	Assessment criterion and evidence
Emergency knowledge	Identifies the applicable disposal procedure for a stated incident and explains the reason for each step	Procedure correctly identified; rationale accurate; applicable regulation cited
Technical response	Operates the relevant equipment and follows the relevant procedure correctly under the abnormal conditions of the scenario	Equipment operated within tolerance; sequence complete; errors recovered without prompting
Situation decision-making	Selects and justifies a defensible course of action when information is incomplete or priorities compete	Decision consistent with available information; alternatives considered; rationale defensible in debrief
Information communication	Transmits accurate, timely and protocol-compliant information to passengers, colleagues and external agencies	Content accurate; timing within drill standard; protocol and phrasing compliant
Teamwork and coordination	Performs an assigned role within a multi-post response and hands over cleanly to the next actor	Role executed; handover complete and unambiguous; shared understanding evidenced
Psychological and adaptive competence	Continues to act effectively while passengers are distressed and time pressure is high	Action maintained under stress; escalation judged appropriately; recovery after error

Competence-based education and the problem of fragmentation

Competence-based education (CBE) derives curriculum objectives from the requirements of real occupational performance, and judges outcomes against those requirements rather than against the coverage of a syllabus (Mulder, 2017). Applied to emergency training, CBE implies that programme outcomes should be written as observable performances under specified conditions, and that industry should participate in defining them, because the standard is set by the workplace rather than by the academy. Biggs and Tang's (2011) principle of constructive alignment supplies the corollary: an intended outcome of integrated performance cannot be assessed by an instrument that samples recall, and a curriculum whose assessment rewards recall will, over time, teach recall regardless of what its objectives declare.

The assessment literature is equally clear about what an authentic instrument requires. Gulikers et al. (2004) specify five dimensions on which assessment tasks approximate professional practice: the task itself, the physical context, the social context, the result or form of output, and the criteria applied. Judged against these dimensions, a written examination of emergency procedures fails on four. It is worth stating why the failure persists, because it is not ignorance of the literature. Authentic assessment is expensive to design, unreliable when raters are untrained, and difficult to defend when results are audited. The administrative response is therefore frequently to retain the vocabulary of competence while measuring what is cheap to measure, a substitution that is invisible in programme documents and visible only in what students can actually do.

Experiential, situated and workplace learning

Kolb's (1984) experiential learning theory explains why the balance between instruction and practice matters pedagogically. Competence develops through a cycle of concrete experience, reflective observation, abstract conceptualisation and active experimentation. A programme that supplies concepts without experience breaks the cycle at its first stage; a programme that supplies drills without structured reflection breaks it at the second.

The cycle is often read as a teaching sequence, which it is not: it is an account of how experience becomes knowledge, and it can therefore be interrupted at any of its four transitions. Schön's (1983) account of reflection-in-action adds a further point: what distinguishes expert practice in uncertain situations is not additional knowledge but the capacity to notice and adjust while acting, which is learned in practice and cannot be supplied by description.

Workplace learning research complicates the assumption that practice, once provided, will do this work on its own. Eraut (2004) shows that most learning at work is informal, reactive and unintended, which means that a curriculum built from designed episodes mis-specifies where competence is actually acquired. Tynjälä (2008) argues that integrating conceptual and experiential knowledge requires deliberate bridging mechanisms, since the two forms of knowledge do not merge of their own accord. Lave and Wenger (1991) push the argument further and, for present purposes, less comfortably: learning is a trajectory of participation in a community of practice, and a learner's position within that community determines what can be learned. Students in a college practise on equipment that is not the operational system, under assessment by teachers rather than by the community whose practice it is, and with consequences that are simulated rather than real. On this account the integration problem is structural, and proximity to an enterprise is not the same thing as participation in it. That distinction is more pessimistic than most partnership literature allows, and it anticipates the third finding of this study.

Simulation-based training: what the evidence supports

Simulation has become a standard component of safety-critical training on the argument that practice in a controlled approximation of the task is preferable to practice on the task itself (Lateef, 2010). The instructional evidence on simulation is substantial, and its central finding is that technology is not the active ingredient. In a systematic review of high-fidelity simulation, Issenberg et al. (2005) identified feedback, repetitive practice, integration into the curriculum and defined outcomes as the features associated with effective learning, rather than the sophistication of the device. Cook et al. (2011), meta-analysing technology-enhanced simulation across health professions, found large effects against no intervention and considerably smaller ones against other forms of instruction, which locates the advantage in practice with feedback rather than in fidelity as such. Hamstra et al. (2014) make the point explicitly: physical fidelity beyond the level required by the learning objective adds cost without proportional benefit, and psychological fidelity, the resemblance of the task to the cognitive demands of practice, is the variable that matters.

Two further results bear directly on the present case. Tannenbaum and Cerasoli (2013), meta-analysing debriefs across team and individual training, report substantial performance gains from structured post-event review, which implies that a drill without debriefing is an expensive way of generating experience that will not be consolidated. The mechanism is feedback: its effect depends on whether it addresses the task, the process or the person, and on whether it tells the learner what to do next (Hattie & Timperley, 2007). Salas et al. (2008) show that team training improves team performance when it identifies team competencies in advance, designs exercises against them and evaluates against the same framework, and that it degrades into collective practice without those steps. Scorgie et al. (2024) extend the evidence to virtual environments in safety training, with the same conditional structure: outcomes depend on instructional design, not on immersion. The implication for a cost-constrained vocational college is therefore not that it needs a simulator. It is that it needs a design in which practice is repeated, varied and debriefed, and in which the scenario, not the device, is the unit of investment.

Industry-education integration and the limits of partnership

International TVET practice treats deep employer involvement as a condition of quality in safety-critical training rather than as an enhancement. Dual-system arrangements embed enterprise practice in the curriculum itself, and competence is assessed jointly against workplace standards (Billett, 2020). Billett (2011) qualifies the claim in a way that matters for policy: workplaces are not automatically educative, since what they afford depends on the organisation of work and on the agency of the learner, so that placing a student in an enterprise guarantees exposure but not learning.

In the Chinese context, cooperation between vocational colleges and URT enterprises is frequently organised through order-class arrangements, in which recruitment, curriculum and placement are negotiated with a single operator. Such arrangements create a channel for deep cooperation but do not guarantee it: the channel is often used for placement, while curriculum content and assessment remain with the college. Sector reports describe the consequences as a shortage of dually qualified teachers, an underdeveloped practical teaching system and a misalignment between educational objectives and enterprise requirements (CAMET, 2023). Andersson and Köpsén (2018) show that vocational teachers maintain occupational currency through continuing activity in their original occupation rather than through initial qualification, which reframes teacher shortage as a workload and organisational problem rather than a hiring problem. What the literature still lacks is evidence of how these conditions appear inside a specific programme, where they can be observed together rather than inferred separately.

Analytical framework

The study uses three lenses, each answering a different question, and Table 2 states them together with the diagnostic question each was used to ask of the data. CBE establishes the standard against which provision is judged, and locates authority for that standard. Experiential and situated learning explains how competence is formed, and therefore what a curriculum must supply. The ADDIE model, comprising analysis, design, development, implementation and evaluation, locates the stage at which a curriculum loses its occupational orientation (Branch, 2009). ADDIE is used here less as a prescription than as a diagnostic sequence: where occupational analysis is thin, objectives drift toward available content; where scenario materials are under-developed, implementation defaults to lecture; where evaluation measures knowledge, the whole cycle reproduces knowledge.

Table 2: Analytical framework and diagnostic questions applied to the data

Lens	Question asked of provision	What a satisfactory answer would require
Competence-based education (Mulder, 2017)	Against what standard are outcomes defined, and who sets it?	Outcomes written as observable performance; employer participation in standard setting
Experiential learning (Kolb, 1984; Schön, 1983)	Does provision supply experience, reflection and repetition?	Supervised practice under varied conditions; structured debriefing after each cycle
Situated learning (Lave & Wenger, 1991)	How close is the learner to participation in operational practice?	Tasks, artefacts and assessors drawn from the operational community
ADDIE (Branch, 2009)	At which design stage does occupational orientation break down?	Occupational analysis whose unit is the incident, not the device

METHODOLOGY

Research design

A qualitative single-case study with embedded units was adopted (Yin, 2018). The case is the emergency rescue skills training provision of one higher vocational college in Chengdu, and the embedded units are its four URT programmes. The case was selected purposively on three criteria: it offers a complete set of URT programmes rather than a single specialisation; it maintains order-class cooperation with local metro operators, so that an industry relationship exists to be examined; and it holds national pilot status for first-aid education, which means emergency content is present rather than absent.

The logic of that selection deserves statement, because single-case research is routinely criticised for what it does not attempt. Flyvbjerg (2006) argues that case selection should be governed by the information a case can yield rather than by its representativeness, and that a critical or typical case supports analytical generalisation where a random sample supports only statistical description. The present case is selected as typical of the better-resourced end of its sector: it has industry partners, a national pilot and recent equipment investment. That choice has a specific consequence for interpretation. A case in which provision is entirely absent would say little about implementation, and a case in which it is exemplary would say little about ordinary conditions; a case that is comparatively well placed, and still reports the constraints described below, yields a conservative estimate of those constraints across the sector.

A single case was also chosen because the research questions concern how and why provision takes the form it does, and because access of the depth required for document analysis and session observation is rarely available across institutions. The design supports analytical rather than statistical generalisation, and transferability to institutions with comparable programme structures.

Sampling and participants

Participants were recruited purposively in three groups whose perspectives differ systematically, so that accounts of provision could be triangulated against accounts of requirement. Twelve students were recruited, three from each of the four programmes, on the criterion that they had completed emergency-relevant coursework or practical training. Four programme leaders or teachers were recruited, one from each programme. Nine enterprise training supervisors were recruited, three from each of three cooperating URT enterprises. Sample size in each group was judged by information power rather than fixed in advance (Malterud et al., 2016), which directs attention to the specificity of the sample, the strength of the dialogue and the analytic strategy rather than to a numerical threshold.

The three groups were not treated as replications of one another but as accounts produced from different positions in the same division of labour, and the analytical value of the design lies in that difference. Students describe what they have been taught and can do; teachers describe what they intend and what constrains them; supervisors describe what the workplace requires and how graduates meet it. Where the three accounts converge, the claim is strong; where they diverge, the divergence is itself the finding.

Table 3: Participants, data sources and contribution to the research questions

Participant group	Number	Selection criterion	Contribution
Students	12 (3 per programme)	Completed emergency-relevant coursework or training	Teaching, drills and assessment as received
Programme leaders and teachers	4 (1 per programme)	Curriculum design or delivery role	Curriculum rationale, resources, assessment decisions
Enterprise training supervisors	9 (3 per enterprise)	Induction or emergency training role	Workplace requirements and graduate performance
Curriculum documents	Programme plans and syllabi	Documents governing the four programmes	Stated objectives, hour allocation, assessment weighting
Observation	Practical sessions	Sessions containing emergency content	Activity structure, equipment use, ratios, debriefing

Data collection

Three sources were used. Semi-structured interviews were conducted with each group, using separate guides: students were asked to describe what they had been taught, what they had practised and how they had been

assessed; teachers were asked about curriculum decisions, resources and constraints; enterprise supervisors were asked what the workplace requires in the first minutes of an incident and how newly recruited graduates perform against that requirement. Interviews were audio-recorded with consent and transcribed verbatim. Curriculum documents, including programme plans and course syllabi, were analysed for stated objectives, hour allocation and assessment weighting. Non-participant observation of practical sessions recorded activity structure, equipment in use, student-to-equipment ratios and the presence or absence of debriefing.

To allow the reader to judge the breadth of the evidence base, the design of data collection is stated here. Interviews were semi-structured, conducted in Chinese, audio-recorded with the participant's consent and transcribed verbatim. Sessions were scheduled for 45 to 60 minutes, and where an account required more time the session was allowed to run beyond that limit; each participant was interviewed once. Twelve students, four programme teachers and nine enterprise training supervisors were interviewed, giving twenty-five interviews in all. Access and role-clarification meetings with programme leaders and enterprise representatives preceded fieldwork; they established permissions, timetables and the division of responsibilities, and they are treated as background rather than as data. Non-participant observation was carried out on site across a single academic semester and covered the three kinds of session through which emergency content is delivered: routine classes that combine explanation with practical work, emergency-specific drill sessions, and Red Cross first-aid certification sessions. Field notes recorded the organisation of the session, the equipment used, the engagement of students with the task and whether a structured debriefing followed the exercise, rather than verbatim talk. The diversity of session types was intended to capture both ordinary delivery and the rarer simulated incidents on which the higher-order dimensions of competence depend.

The three sources were treated as complementary rather than substitutable: documents state intent, interviews report experience, and observation records what occurs. Divergence between them was treated as data rather than as error, and is reported where it bears on the findings. This decision reflects a methodological judgement about the object of study. The study is not attempting to measure the frequency of particular teaching behaviours, for which observation alone would be required; it is attempting to characterise the design of provision, which is visible in what documents require, what teachers intend and what sessions are organised to make possible.

Data analysis and trustworthiness

Interview and observational data were analysed using Braun and Clarke's (2006) six-phase thematic analysis. Coding was inductive at the first stage, generating codes close to participants' accounts; the second stage organised those codes against the study's theoretical framework, so that the final themes describe both what participants reported and how it is explained. Document data were analysed through a content matrix aligned to the same framework, which allowed intent and practice to be compared directly. Coding was managed so that a theme was retained only where it was supported by at least two of the three data sources, a rule that deliberately privileges convergence over volume of testimony.

To make the analytic procedure auditable, the following details are recorded. The first author carried out the whole coding procedure in both cycles, and coding was managed in qualitative analysis software. Following Malterud et al. (2016), information power rather than numerical saturation governed sampling and the stopping rule, assessed against five dimensions: the aim, which was narrow and specific, a single case with four embedded programmes; the population, small but purposively chosen; the specificity of the sample, students who had completed emergency-relevant coursework, programme leaders responsible for curriculum and enterprise supervisors with induction or emergency-training roles; the quality of the dialogue, which the field notes record as sustained and specific; and the analytic strategy, a theory-informed thematic analysis. Data collection was closed when further interviews and observations yielded elaboration of existing codes rather than new ones, and every analytic step was documented in reflexive memos. No second independent coder was available for this study; credibility was therefore secured through triangulation across the three data sources, member checking of transcript summaries with selected participants and the audit trail, and the absence of parallel coding is recorded as a limitation in the concluding section.

Trustworthiness was addressed on the four criteria proposed by Lincoln and Guba (1985), and reported following the transparency standards set out by Nowell et al. (2017). Credibility was supported by triangulation across the three data sources and by member checking of transcript summaries with participants. Transferability was supported by a thick description of the case, its programmes and its industry relationships. Dependability was supported by an audit trail linking codes to themes and themes to data. Confirmability was supported by reflexive memo-writing, in which the first author recorded how prior professional experience in vocational education shaped the reading of the data, particularly the inclination to read constraint as institutional rather than individual failure.

Ethical considerations

Approval was obtained from the researcher's institution and from the participating college. All participants received an information sheet, gave written informed consent, and were told that participation was voluntary and could be withdrawn. Names, programme identifiers and enterprise identifiers were replaced with codes during transcription, and verbatim material was checked for identifying detail before use.

Boundaries of the design

Two features of the design bound what the findings can claim, and both are stated here rather than deferred to the limitations. Interviews report what participants believe and are willing to say, and testimony about practice is not practice. Observation records what occurred in sessions to which access was granted, and the presence of an observer may alter what occurs. Neither limitation is fatal to the argument advanced here, because the claims concern the design of provision rather than the frequency of behaviour, and design is visible in documents and in the arrangement of sessions as well as in testimony. A third boundary is temporal: data were generated at one point in time, so the account describes provision as it stood rather than how it is changing.

RESULTS

Provision of emergency rescue content

Emergency rescue content is present in all four programmes, but it is concentrated rather than distributed. The principal vehicle is a core course on operational safety and emergency handling, supported by first-aid modules delivered under the college's pilot status and by induction content within order-class arrangements. Documents show the stated objectives in competence language: students are expected to identify hazards, execute standard disposal procedures and cooperate in emergency response. The internal structure of the provision, however, is organised by knowledge domain rather than by incident type, and the six dimensions of emergency competence identified in the literature appear unevenly. Knowledge and technical response are addressed explicitly; decision-making, communication and psychological adaptation are named in objectives but rarely developed as separate learning outcomes.

Concentration has a consequence that is easy to overlook in a programme-level reading. Because the content sits in one course, its quality is decided by one course team, and the remaining programmes inherit whatever that team can sustain. The arrangement is efficient in the short term and fragile in the medium term: it produces a single point of design capacity on which four programmes depend, and it makes the emergency curriculum difficult to revise without revising the course that hosts it.

Physical resources exist and are used. Simulated station environments, vehicle training equipment and first-aid kits support practical sessions, and the first-aid route leads to a recognised certificate that students value independently of the programme. Observation confirmed that sessions are organised and supervised. The first-aid route deserves separate comment, because it is the one component of provision that is externally specified, externally examined and recognised by employers, and it demonstrates within the same institution what competence-based assessment requires: a defined performance standard, an assessor who is not the class teacher, and a consequence attached to failure. Its scope, however, is basic life support. It does not reach the dimensions of decision-making, communication and coordination on which the argument of this paper turns, and its success

is therefore evidence of what the institution can do when a standard is externally imposed rather than evidence that its own standards are performance-based.

A theory-practice gap in competence development

The first challenge concerns the balance between instruction and practice. Across the documents examined, the greatest share of allocated hours falls to classroom delivery, with practical hours concentrated in a small number of modules. Teachers described the reason in terms of coverage: the syllabus is extensive and the examination timetable fixed, so practical sessions absorb the time that remains. Students described the consequence in terms of confidence: procedures are known as sequences on paper, and the experience of executing them under pressure is limited.

Documents give the allocation in hours. The compulsory course Urban Rail Transit Emergency Handling carries 72 contact hours, of which 48 are classroom-based and 24 are practical or drill-based; the module Emergency First Aid Knowledge and Skills carries 36 hours, divided into 12 hours of classroom instruction and 24 hours of supervised practice (Chengdu Industry and Trade College, 2026). Emergency content also appears inside other professional courses and within order-class induction, but without separately allocated hours. These two modules are delivered with comparable allocations across the four programmes, so that the emergency course is weighted two to one towards the classroom while the first-aid module is weighted two to one towards practice. The figures qualify the qualitative account above rather than standing beside it: the course that carries the knowledge of procedures is predominantly lectured, the module that teaches hands-on technique is predominantly practised, and the scenario-based decisions that would connect the two have no allocated hours of their own.

The allocation is not arbitrary, and its internal logic is worth setting out because it explains why exhortation has not changed it. The allocation follows the examination; the examination follows the syllabus; the syllabus follows the textbook. Each step is individually defensible and each is defended by someone. The sequence as a whole, however, produces a curriculum in which practice is what remains after coverage has been satisfied, and in which the residual is distributed to whatever module can accommodate it. This is a design failure rather than a failure of commitment: the same teachers who described the constraint also described wanting more drill time.

Observation sharpened the point. Sessions were typically organised around equipment operation and correct procedure, with little variation in incident conditions and limited debriefing. Where an incident was simulated, the scenario rarely required students to decide between competing priorities, to communicate with another department, or to manage passengers who do not comply. These are precisely the conditions under which the higher-order dimensions of competence are exercised, and their absence narrows what practice can develop. The point is not that the scenarios were unrealistic; it is that their realism was technical rather than psychological, reproducing the equipment with some accuracy while omitting the informational and social conditions that make an incident difficult (Hamstra et al., 2014).

Students described the effect of this balance on their own readiness, and the description recurred across programmes. One student (S07) distinguished between being able to recite the sequence and being able to run it, and described the practical component as something demonstrated to the class rather than rehearsed by it; a second student used the same distinction in different words, describing a single demonstration as all the practice the module allowed. The accounts show how the allocation of hours is experienced by learners: the sequence is known on paper, but opportunities to execute it under conditions that resemble an incident are limited.

Assessment reinforces the same pattern. Where the division between the written paper and the practical assessment is not prescribed, the written paper remains the instrument that decides the mark, and practical assessment is used for a smaller number of operationally defined tasks. Enterprise supervisors reported that newly recruited graduates can state procedures accurately and are less reliable when an incident deviates from the routine case. That is not evidence that graduates are unprepared in general; it is evidence that the assessment instrument measures the dimension it is designed to measure, and leaves the others unassessed. Students, for their part, did not describe their programmes as deficient. They valued the first-aid certificate and the order-class placement, and treated the written examination as the authoritative measure of competence, which is itself a finding: assessment does not only measure what students know, it teaches them what counts as knowing.

The same substitution appeared in how students described their own competence. Asked what it meant to be prepared for an incident, one student (S11) answered in terms of the result on the paper rather than the way the drill had gone – an account that treats the examination as the evidence of readiness rather than the exercise itself.

For clarity, the weighting is stated. The two modules are assessed differently, and neither institutional document prescribes a fixed percentage division between the written paper and the practical assessment. In Urban Rail Transit Emergency Handling the final mark combines a written theoretical examination with a practical operational assessment, and the proportion between them is set by the course teacher within the syllabus; the first-aid module is assessed externally through Red Cross certification, which is reported separately and is not incorporated into the module mark (Chengdu Industry and Trade College, 2026). The absence of a prescribed weighting is itself part of the finding: the assessment dimension of the gap is not a deliberate allocation of marks to recall over performance but an unspecified arrangement in which the written paper, being the instrument that can be marked consistently across classes, ends by carrying the decisive weight.

Constraints of the current training model

The second challenge concerns the capacity of existing arrangements to support the practice that emergency competence requires. Two constraints recurred. The first is the orientation of facilities. Investment has followed the equipment-intensive parts of the curriculum, which is consistent with programme priorities but leaves the scenarios most relevant to emergency response, namely passenger evacuation, station crowd management and multi-agency joint rescue, thinly resourced. Where these are taught, they are taught as procedure rather than as simulation.

The pattern of investment is intelligible once its decision criteria are made explicit. Equipment is visible, procurable, countable and auditable; a scenario is none of these. A vehicle training rig can be inspected, photographed and entered in an asset register, while a well-designed evacuation exercise leaves behind nothing but a debriefing note. Where capital decisions are justified by what can be shown, provision will drift toward hardware irrespective of its instructional yield, and the drift will be invisible in the documents that record it, because each individual decision is defensible on its own terms.

The second constraint is teaching capacity. Several teachers described their operational knowledge as dating from earlier industrial employment, and reported few recent opportunities to observe current front-line practice. Their pedagogical competence was not in question; what was limited was currency of the conditions against which they teach. This is the shortage of dually qualified teachers that sector reports identify nationally (CAMET, 2023), appearing here as an individual and specific constraint rather than a statistical one. It also illustrates why the shortage is resistant to policy: the remedy is not a qualification but recurrent access to an operating environment, and access of that kind has to be organised, resourced and scheduled (Andersson & Köpsén, 2018).

The constraint of teacher currency was described by the teachers themselves. One programme leader (T03) traced his own operational knowledge to earlier industrial employment and dated his most recent access to a live operating environment from that period; a second teacher (T06) made the same point, describing the conditions against which he teaches as ones he observes only at second hand. Their accounts locate the shortage of dually qualified teachers in the absence of recurrent access to operations rather than in a lack of formal qualification, and they support the reading of the mechanism offered below.

Shallow industry-education integration

The third challenge concerns the depth of enterprise involvement. Order-class arrangements provide a structured relationship with operators, and enterprise supervisors participate in induction, placement supervision and occasional guest sessions. Their participation in curriculum design, drill facilitation and competence assessment, however, is limited. Enterprise requirements are communicated mainly through informal channels and at the point of recruitment, and are only partly converted into teaching objectives and assessment criteria.

The mechanism is contractual rather than attitudinal. The order-class agreement specifies numbers of places, selection arrangements and employment terms; it does not specify module content, and where it is silent, curriculum reverts to the college by default. The consequence is a translation gap rather than an absence of cooperation. Enterprises know what the first minutes of an incident demand, and colleges know what a programme can deliver within its constraints, but the two accounts are rarely reconciled in a document that governs teaching. Enterprise supervisors described requirements in terms of judgement and communication under stress; programme documents describe them in terms of correct procedure. Both are right, and the difference between them is what a jointly designed curriculum would resolve.

It is worth recording what the supervisors did not say. None described the relationship as unwelcome, and several described willingness to contribute to curriculum and assessment work if it could be organised within their operating duties. The obstacle reported was not disposition but the absence of a forum, a schedule and a division of labour through which contribution could occur. That distinction matters for the recommendations that follow, since it locates the problem in the design of cooperation rather than in the goodwill of its parties.

Enterprise supervisors made the workplace requirement explicit in interview. One supervisor responsible for new-employee induction (E05) described the demand of the first minutes as a demand for judgement and communication under conditions in which the rule book is necessarily incomplete; a second (E08) described the gap in graduate performance in the same terms, as an ability to state a procedure without an ability to decide within it. These accounts place the judgement and communication demands of the first minutes in the testimony of those who employ graduates, and they anchor the translation gap described above in workplace evidence rather than in inference.

Assets on which improvement can build

Provision is not deficient across the board, and the findings would be misread if they suggested otherwise. Three assets are available. Institutional policy supports safety education and has been willing to invest in training equipment. The first-aid certification route provides an existing, externally validated component that demonstrates what competence-based assessment looks like in practice. Order-class partnerships provide the relationship through which deeper cooperation can be negotiated. These assets matter because the improvements implied by the findings do not require new structures, but different use of existing ones.

DISCUSSION

The three challenges identified above are usually discussed in this literature as separate deficits, each with its own remedy: more practical hours, better facilities, deeper enterprise involvement. The evidence from this case does not support that decomposition. The three are manifestations of one failure, namely a curriculum organised by subject content rather than by the occupational performance the sector requires, and they persist because that organisation is reproduced by mechanisms that operate independently of the intentions of the people inside it.

Why the gap persists: five mechanisms

The first mechanism is the audit logic of assessment. Written examination is cheap, reliable, comparable across cohorts and defensible to an inspectorate; performance assessment is expensive, rater-dependent and hard to defend when results are contested (Gulikers et al., 2004). An institution under audit pressure will therefore gravitate toward the instrument it can defend, and the fact that the instrument samples the wrong construct will not appear in any report.

The second is the visibility of capital. Facilities follow equipment because equipment can be counted, whereas scenario design cannot, and investment criteria reward what can be shown. The consequence is a systematic under-investment in the least visible input, which happens also to be the one most strongly associated with learning gains (Issenberg et al., 2005).

The third is the formation of teachers. Occupational currency decays without recurrent practice, and recurrent practice requires an organisational arrangement that no individual teacher can create alone (Andersson &

Köpsén, 2018). A college can require certification; it cannot, by requiring it, restore the experience that certification was meant to represent.

The fourth is the contractual scope of partnership. Order-class agreements are agreements about employment, and curriculum is left to the college by default rather than by decision. The partnership is therefore deep where the contract reaches and shallow where it does not, which is precisely the pattern the third finding describes. This is a more precise diagnosis than the familiar claim that enterprises are unwilling to engage, and it points to a different remedy: the negotiation of a small number of curriculum obligations within an instrument that already exists.

The fifth is the absence of a feedback loop from operations to curriculum. Near misses, drill findings and incident reviews generate exactly the material from which scenarios should be built, and that material circulates within the operator rather than returning to the college. Reason's (1997) account of latent conditions implies that an organisation which does not route operational learning back into training is accumulating the conditions for its next incident, and the same logic applies to the college that supplies its staff.

Re-reading the three challenges through the framework

Read through the experiential lens, the theory-practice gap is a break in the learning cycle at its second transition. Where practice is scarce and debriefing limited, students accumulate concrete experience without the reflective observation that converts it into transferable capability, and instruction supplies conceptualisation without the experience it conceptualises (Kolb, 1984). The result is knowledge that is available for examination and less available under pressure, which is the pattern enterprise supervisors described. That diagnosis also explains why adding hours alone would not close the gap: additional practice without structured review reproduces experience without consolidation (Tannenbaum & Cerasoli, 2013).

Read through ADDIE, the training-model constraint appears as a weakness at the analysis and development stages. Facilities oriented toward equipment reflect an earlier occupational analysis, in which operational competence meant operating a device correctly. Emergency competence requires an analysis whose unit of performance is an incident, and until that analysis is revised, resource decisions will reproduce the earlier curriculum whatever the objectives declare (Branch, 2009). The same reading applies to assessment: an evaluation stage that measures knowledge will feed back into a design stage that specifies knowledge, and the cycle will stabilise around the measurable.

Read through CBE, the integration challenge is the design-stage problem behind both. The standard against which outcomes are judged is set by occupational performance, and the authority to define it is shared with employers (Mulder, 2017). Where enterprises are involved at placement but not at design, the standard effectively defaults to what the college can specify alone, and assessment reverts to what the college can measure reliably without enterprise participation. Cooperation then becomes a channel used for recruitment rather than a mechanism for curriculum. The situated-learning lens explains why this default is so stable: students are not peripheral participants in the operator's community of practice, and no amount of goodwill between the parties changes their position unless tasks, artefacts and assessors are drawn from that community (Lave & Wenger, 1991).

Comparison with dual-system provision

Compared with dual-system arrangements in which emergency units are embedded in occupational qualifications and assessed jointly (Billett, 2020), the case shows a relationship with the right partners and the wrong depth. The comparison is not a judgement of quality: dual-system provision rests on institutional conditions, including employer associations with statutory training responsibilities and sectoral qualification frameworks, that no single college can create. Nor is the comparison a recommendation of borrowing. What the case suggests is narrower and more practicable: depth can be increased within existing arrangements, through joint design of a small number of emergency modules, joint facilitation of drills, and joint assessment of performance against criteria that enterprises help to write.

Comparison with emergency training research in other safety-critical sectors

Beyond vocational education, the strongest empirical evidence on emergency training comes from other safety-critical sectors, and it helps to locate the present findings. In healthcare, systematic reviews of simulation-based training consistently show that what produces learning is not the fidelity of the equipment but feedback, repetitive practice, integration into the curriculum and explicitly defined outcomes (Issenberg et al., 2005; Cook et al., 2011). The same reviews caution that physical fidelity beyond the level required by the learning objective adds cost without proportionate benefit, and that psychological fidelity, the resemblance of the task to the cognitive demands of practice, is the decisive variable (Hamstra et al., 2014). Findings on debriefing reinforce the point: structured post-event review produces substantial performance gains across team and individual training (Tannenbaum & Cerasoli, 2013), and team training improves performance only when team competencies are specified in advance, exercises are designed against them and evaluation uses the same framework (Salas et al., 2008). Evidence from virtual environments for safety training carries the same conditional structure, with outcomes depending on instructional design rather than on immersion (Scorgie et al., 2024). Read together, this literature treats the scenario, the feedback and the debrief as the unit of investment, and treats equipment as a support rather than the object of training. That is precisely the inversion the present case has not made.

The empirical record from these sectors is not only a record of effect sizes. Two studies speak directly to the institutional mechanisms identified in this case. In maritime education, Emad and Roth (2008) followed a competency-based training system through its regulations, syllabi, examination papers and classroom practice, and found that the assessment regime had quietly redefined the object of training: candidates were prepared to pass a competency examination rather than to act competently on board, and the curriculum was redirected not by its content but by what was measured. That is the substitution reported in the present case, arrived at independently in another industry and another national system. In aviation, Salas et al. (2006) reviewed the evaluations of crew resource management training and found consistent positive effects on attitudes and learning, weaker and less consistent evidence that these gains transferred to behaviour in operations, and attributed the shortfall to the absence of recurrent practice, reinforcement and organisational conditions that sustain the trained behaviour. The parallel is structural rather than superficial: a training intervention that succeeds within its own assessment can still fail at the point of transfer unless the surrounding organisation keeps it alive.

The comparison must, however, be bounded. These studies examine qualified or qualifying professionals in sectors with established qualification frameworks, high training budgets and mature regulatory regimes; the present case concerns vocational students in a resource-constrained college where the emergency component is one course among four programmes. The learners, the standard of performance and the institutional consequences of failure therefore differ, and effects reported in healthcare or aviation cannot be transferred as effect sizes to a Chinese higher vocational college. What transfers is the design principle rather than the result: practice must be repeated, varied and debriefed, and fidelity must be judged against the cognitive demands of the incident rather than the physical accuracy of the equipment. The comparison also exposes what this case cannot show. Because the cited studies measure learning outcomes under controlled conditions, they cannot speak to the institutional mechanisms, the audit logic of assessment or the contractual limits of partnership that the present case identifies as the sources of the gap. Those mechanisms are visible only in an account of provision such as this one, which is why the comparison is used here to sharpen the diagnosis rather than to supply it.

These studies also mark the limits of the comparison. Emad and Roth's account is a single institution in one jurisdiction, and Salas et al.'s conclusions rest on primary evaluations whose quality and heterogeneity constrain what can be aggregated; neither design can establish how widely the mechanisms they describe are distributed across institutions. They are cited here as independent corroboration of a mechanism, not as evidence of its prevalence, and prevalence is precisely what the comparative and longitudinal designs proposed in the conclusion would begin to establish.

Leverage points and implications

The argument so far implies that reform should concentrate on the points where the five mechanisms are weakest, rather than on the symptoms they produce. Table 4 states the resulting priorities against the challenge each

addresses, the mechanism it interrupts and an observable indicator of success. Two features of the table are deliberate. First, every action lies within the authority of a college or of a college in negotiation with an existing partner, so that the recommendations do not depend on national reform for their implementation. Second, every action has an indicator that can be observed without new measurement apparatus, because recommendations that require a research project to evaluate are rarely implemented twice.

Table 4: Challenges, mechanisms and leverage points

Challenge	Mechanism sustaining it	Leverage point	Indicator of success
Theory-practice gap	Assessment audit logic	Rewrite outcomes as performance and assess them with enterprise-written criteria	Practical assessment carries equal weight in at least one core module
Theory-practice gap	Unstructured drill cycles	Mandatory structured debriefing within every scenario session	Debrief record exists for each drill, with student reflection logged
Training-model constraint	Visibility of capital in investment decisions	Treat scenario design as a budget line rather than an activity	Scenario library funded and versioned alongside equipment
Training-model constraint	Decay of teacher occupational currency	Scheduled front-line attachment for emergency-module teachers	Recurrent attachment agreed in the partnership instrument
Shallow integration	Contractual scope limited to employment	Add curriculum obligations to the order-class agreement	Jointly signed module specification for at least two modules
All three	No operational feedback loop	Route anonymised drill and near-miss findings into scenario design	Cases in the scenario library traceable to operational events

Contribution, transferability and limitations

The study's analytic contribution is to replace a list of deficits with a mechanism account. That replacement has consequences for how the field's standard recommendations should be read. Calls for more practical hours, more simulation and deeper enterprise involvement are not wrong, but they are underspecified: each names a quantity where the evidence points to a design condition. Practice helps when it is varied and debriefed; simulation helps when its fidelity is psychological rather than physical (Hamstra et al., 2014); enterprise involvement helps when it reaches design and assessment rather than placement alone. The distinction matters in a sector where resources are fixed, because it directs scarce capacity toward design rather than toward procurement.

Transferability rests on the similarity of programme structure rather than on statistical representativeness. Colleges that host a complete set of URT programmes, maintain order-class relationships and hold some externally validated safety certification will recognise the configuration described here; colleges without those features will not, and the mechanisms identified may operate differently where industry partnership is weak for reasons of geography rather than design. The claim advanced is therefore conditional: where the mechanisms are present, the leverage points in Table 4 are available, and where they are absent the analysis does not apply.

Four limitations bound these conclusions. The study is a single case, and its findings describe one institution in one sector; they are transferable to comparable settings, not generalisable across the system. Coding was carried out by a single researcher without a second independent coder, so a unilateral reading of the material cannot be excluded, although triangulation across the data sources, member checking and the audit trail constrain it. Data

were generated at one point in time, so the account captures provision as it stood rather than how it is changing. And the study did not follow graduates into employment, so the relationship between curriculum design and later operational performance is inferred from supervisor accounts rather than observed. That last limitation is the most consequential, and it defines the next stage of the work.

CONCLUSION

This paper examined how emergency rescue skills training is implemented in the four URT programmes of one Chinese higher vocational college. Provision exists, is resourced, and is organised around a core course supported by first-aid certification and industry partnership. Its limits are structural rather than incidental: practical time is subordinate to coverage, facilities follow equipment rather than incidents, assessment measures knowledge more reliably than performance, and enterprise involvement stops short of curriculum design and competence assessment. Each of these limits can be traced to the same cause, namely a curriculum organised by subject content rather than by the occupational performance the sector requires, and each is sustained by a mechanism that operates whether or not anyone intends it.

Four directions follow. Emergency modules should be rebuilt around the six dimensions of emergency competence, with outcomes written as observable performance under stated conditions. Scenario simulation should be expanded and concentrated on the conditions that carry the higher-order dimensions: time pressure, incomplete information, cross-department communication and non-compliant passengers. Assessment should combine knowledge testing with performance assessment against criteria that enterprises help to define, and every drill should end in a structured debrief. Teacher currency should be maintained through scheduled front-line attachment, not only through certification. Finally, enterprise involvement should be moved upstream, from placement to joint design and joint assessment of a defined set of modules, written into the instrument that already governs the partnership.

For a sector whose incidents are rare and whose consequences are concentrated, the value of this reconstruction lies less in the average competence of graduates than in the reliability of the weakest performance on the worst day. That formulation has a methodological consequence. Average competence can be estimated from examination results; the reliability of the weakest performance cannot, and requires observation of graduates under conditions the college does not control. Future research should proceed in two directions, both of which address the limits of the present design. The first is multi-institutional comparative empirical research. A comparative study across colleges that differ in resources, programme structure and the form of enterprise partnership would test which of the mechanisms identified here are general to the sector and which are specific to this case, and would establish the conditions under which the proposed leverage points travel. The second is longitudinal research that follows graduates into their first years of operational duty and measures their emergency response performance directly, rather than inferring it from supervisor accounts. Tracking the same cohorts from college into employment would show whether competence developed in college survives contact with operational practice, whether the reliability of the weakest performance holds under real incident conditions, and how much of that performance is attributable to particular features of pre-service training. Together, comparative and longitudinal designs would replace the analytical generalisation offered here with empirical evidence about variation between institutions and change over time.

ACKNOWLEDGEMENTS

The authors thank the participating college, its programme teams and the three cooperating enterprises for access to documents, sessions and staff time, and the students and supervisors who gave their time to the interviews.

REFERENCES

1. Andersson, P., & Köpsén, S. (2018). Maintaining competence in the initial occupation: Activities among vocational teachers. Vocations and Learning, 11(2), 317–344. https://doi.org/10.1007/s12186-017-9192-9
2. Biggs, J., & Tang, C. (2011). Teaching for quality learning at university (4th ed.). Open University Press.

3. Billett, S. (2011). Vocational education: Purposes, traditions and prospects. Springer. https://doi.org/10.1007/978-94-007-1954-5
4. Billett, S. (2020). Learning in the workplace: Strategies for effective practice. Routledge. https://doi.org/10.4324/9781003116318
5. Branch, R. M. (2009). Instructional design: The ADDIE approach. Springer. https://doi.org/10.1007/978-0-387-09506-6
6. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
7. Chengdu Industry and Trade College. (2026). Talent training programme for urban rail transit operation management [Unpublished internal document]. Chengdu Industry and Trade College.
8. China Association of Metros. (2023). Talent cultivation plan for urban rail transit during the 14th Five-Year Plan period. https://www.camet.org.cn/rcpy/gh/12425.shtml
9. China Association of Metros. (2026). Urban rail transit 2025 annual statistics and analysis report. https://www.camet.org.cn/xytj/tjxx/789653532090437.shtml
10. Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J., & Hamstra, S. J. (2011). Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. JAMA, 306(9), 978–988. https://doi.org/10.1001/jama.2011.1234
11. Emad, G., & Roth, W. M. (2008). Contradictions in the practices of training for and assessment of competency: A case study from the maritime domain. Education + Training, 50(3), 260–272. https://doi.org/10.1108/00400910810874026
12. Eraut, M. (2004). Informal learning in the workplace. Studies in Continuing Education, 26(2), 247–273. https://doi.org/10.1080/158037042000225245
13. Flyvbjerg, B. (2006). Five misunderstandings about case-study research. Qualitative Inquiry, 12(2), 219–245. https://doi.org/10.1177/1077800405284363
14. Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. Educational Technology Research and Development, 52(3), 67–86. https://doi.org/10.1007/BF02504676
15. Hamstra, S. J., Brydges, R., Hatala, R., Zendejas, B., & Cook, D. A. (2014). Reconsidering fidelity in simulation-based training. Academic Medicine, 89(3), 387–392. https://doi.org/10.1097/ACM.0000000000000130
16. Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112. https://doi.org/10.3102/003465430298487
17. Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Lee Gordon, D., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. Medical Teacher, 27(1), 10–28. https://doi.org/10.1080/01421590500046924
18. Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. Prentice-Hall.
19. Lateef, F. (2010). Simulation-based learning: Just like the real thing. Journal of Emergencies, Trauma, and Shock, 3(4), 348–352. https://doi.org/10.4103/0974-2700.70743
20. Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge University Press. https://doi.org/10.1017/CBO9780511815355
21. Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. SAGE Publications.
22. Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. Qualitative Health Research, 26(13), 1753–1760. https://doi.org/10.1177/1049732315617444
23. Ministry of Transport of the People's Republic of China. (2018). Provisions on the administration of urban rail transit operations (Order No. 8). https://xxgk.mot.gov.cn/2020/jigou/fgs/202006/t20200623_3307960.html
24. Mulder, M. (2017). Competence theory and research: An introduction. In M. Mulder (Ed.), Competence-based vocational and professional education: Bridging the worlds of work and education (pp. 3–34). Springer. https://doi.org/10.1007/978-3-319-41713-4_1
25. Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. International Journal of Qualitative Methods, 16(1), 1–13. https://doi.org/10.1177/1609406917733847

26. Reason, J. (1997). Managing the risks of organizational accidents. Ashgate.
27. Salas, E., DiazGranados, D., Weaver, S. J., & King, H. (2008). Does team training work? Principles for health care. Academic Emergency Medicine, 15(11), 1002–1009. https://doi.org/10.1111/j.1553-2712.2008.00254.x
28. Salas, E., Wilson, K. A., Burke, C. S., & Wightman, D. C. (2006). Does crew resource management training work? An update, an extension, and some critical needs. Human Factors, 48(2), 392–412. https://doi.org/10.1518/001872006777724444
29. Schön, D. A. (1983). The reflective practitioner: How professionals think in action. Basic Books. https://doi.org/10.4324/9781315237473
30. Scorgie, D., Feng, Z., Paes, D., Parisi, F., Yiu, T. W., & Lovreglio, R. (2024). Virtual reality for safety training: A systematic literature review and meta-analysis. Safety Science, 171, Article 106372. https://doi.org/10.1016/j.ssci.2023.106372
31. State Council of the People's Republic of China. (2019). National implementation plan for the reform of vocational education [国家职业教育改革实施方案]. http://www.gov.cn/zhengce/content/2019-02/13/content_5365341.htm
32. Tannenbaum, S. I., & Cerasoli, C. P. (2013). Do team and individual debriefs enhance performance? A meta-analysis. Human Factors, 55(1), 231–245. https://doi.org/10.1177/0018720812448394
33. Tynjälä, P. (2008). Perspectives into learning at the workplace. Educational Research Review, 3(2), 130–154. https://doi.org/10.1016/j.edurev.2007.12.001
34. Weick, K. E., & Sutcliffe, K. M. (2015). Managing the unexpected: Sustained performance in a complex world (3rd ed.). Wiley.
35. Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). SAGE Publications.