



Meta-Skills as Managed Outcomes: A Delphi-Validated Implementation System for Chinese Private HEIs

Chi Che

Educational Leadership Department, International College, Payap University, Chiang Mai, Thailand

*Corresponding Author, e-mail: chi_c@payap.ac.th

Abstract

Chinese private higher education institutions are under growing pressure to demonstrate graduates' meta-skills, yet implementation often stalls because curriculum, pedagogy, assessment, governance, and quality assurance are not sufficiently aligned. This study developed and validated an institution-level implementation playbook to support the systematic integration of meta-skills in Chinese private HEIs. Building on an earlier design-thinking-based framework, the study used a two-round modified Delphi process with 15 experts from private higher education institutions in Sichuan Province. Expert judgments were used to evaluate suitability, feasibility, and clarity, together with content validity indices and qualitative feedback, to refine five implementation domains: governance and decision rights, staged rollout, quality assurance embedding, evidence routines and key performance indicators (KPIs), and risk controls. Round 2 results showed stronger consensus, improved clarity, and high perceived suitability and feasibility across domains. Experts particularly emphasized the importance of explicit decision rights, protected workload, moderation routines, curriculum mapping, and minimum evidence standards for sustainable implementation. The findings indicate that meta-skills become scalable and auditable only when supported by coordinated institutional routines rather than isolated innovation efforts. The validated playbook offers a practical and transferable pathway for Chinese private HEIs operating under comparable governance and resource conditions.

Keywords: Meta-skills, Implementation Governance, Curriculum Integration, Quality Assurance, Delphi Validation

Introduction

Meta-skills as institutionally accountable learning outcomes

Higher education institutions (HEIs) are increasingly judged not only by disciplinary attainment but also by whether graduates can demonstrate transferable capabilities under uncertainty, technological change, and complex social demands (Otermans et al., 2025; Tuononen et al., 2022; Brown et al., 2021). These capabilities are widely discussed as meta-skills: higher-order, cross-context competencies that shape how individuals learn, collaborate, create, and exercise judgment across academic and professional settings (Prasittichok & Klaykaew, 2022; Tuononen et al., 2022). Rather than peripheral “soft skills,” meta-skills are increasingly understood as explicit learning outcomes that should be progressively developed and evidenced through observable performance, assessment, and feedback over time (Vlachopoulos & Makri, 2024; Sokhanvar et al., 2021; Banihashem et al., 2022). Translating meta-skills into assessable outcomes requires learning and assessment practices that accommodate diverse learners and make complex capabilities visible. Personalized adaptive learning can respond to individual learning needs (du Plooy et al., 2024), while flexible assessment can strengthen learner agency and connections with future practice (Elkington & Chesterton, 2025). Authentic assessment also encompasses relational engagement (Gravett, 2025), socially meaningful participation (McArthur, 2023), and opportunities for students to contribute knowledge (Nieminen et al., 2025). Within digitally mediated learning, authentic tasks offer opportunities to develop digital creativity (Janse van Rensburg et al., 2022), supported by purposeful integration of digital practices into assessment design (Nieminen et al., 2023). However, evidence from vocational education reveals gaps between authentic assessment principles and enacted practices (Villarroel et al., 2024), while broader research emphasizes deliberate coordination of assessment tasks, criteria, feedback, agency, and collaboration (Zhan et al., 2025). Together, these perspectives inform the present playbook’s integration of curriculum, pedagogy, assessment, and quality assurance to support meaningful and demonstrable meta-skills development.

In this study, meta-skills are organized as a four-domain capability bundle: adaptive expertise or learning agility, relational dynamics, creative agility, and strategic synthesis. This structure is not presented as a direct borrowing from a single prior source, but as a context-grounded synthesis informed by recent literature on graduate capabilities and refined through earlier framework-development work in Sichuan private HEIs (Tuononen et al., 2022; Brown et al., 2021; Che, 2024). The four domains were selected because, together, they capture the forms of capability most relevant to

contemporary higher education and professional practice: learning under novelty, working effectively with others, generating and refining ideas, and integrating perspectives for strategic judgment.

Despite growing policy and institutional attention to such outcomes, a persistent management problem remains. Many HEIs now reference graduate attributes similar to meta-skills, yet these outcomes often remain aspirational rather than institutionally accountable. Program documents may endorse them, but curriculum mapping, pedagogical support, assessment governance, and quality assurance (QA) routines frequently continue to prioritize content coverage and conventional grading practices. Under such conditions, meta-skills are difficult to compare across programs, difficult to audit, and difficult to improve systematically (Krooi et al., 2024; Sorour & Atkins, 2024). Needs-assessment evidence also indicates measurable gaps between students' current and desired meta-skill states, suggesting that meta-skills development cannot be left to isolated classroom innovation alone; it requires coordinated institutional mechanisms and credible evidence loops (Prasittichok & Klaykaew, 2022; Chen et al., 2023).

The implementation problem in Chinese private higher education

This challenge is especially visible in Chinese private HEIs. As market-oriented institutions operating under enrollment competition and intensified employability expectations, they face strong pressure to demonstrate graduate readiness while often working under constraints in staffing capacity, workload flexibility, assessment moderation, and data infrastructure (Brown et al., 2021; Mok & Han, 2017; Xu, 2021). In such settings, reforms are vulnerable when they depend on informal coordination, short-term enthusiasm, or isolated champions rather than stable institutional routines (Alzahmi et al., 2025; Harris et al., 2022; Nadeem, 2024).

In practice, the absence of clear decision authority, protected time for redesign and moderation, standardized assessment expectations, and minimum evidence requirements can cause otherwise promising reforms to fragment across programs, drift into symbolic compliance, or collapse under workload strain (Krooi et al., 2024; Sorour & Atkins, 2024; Banihashem et al., 2022). The central issue, therefore, is not whether meta-skills matter, but whether institutions can reliably coordinate curriculum, pedagogy, assessment, and QA so that meta-skills become measurable and improvable at scale. For Chinese private HEIs, this creates a stringent test of feasibility and sustainability: any workable model must operate under resource constraints, remain resilient to personnel turnover, and fit within routine QA cycles rather than function as a parallel initiative (Mok & Han, 2017; Xu, 2021; Krooi et al., 2024, Che et al., 2026a, 2026b).

From innovation design to institutionalization: conceptual framing and research questions

Design thinking (DT) has gained prominence in education because of its iterative, stakeholder-centered logic and its capacity to generate learner-centered innovations (Hasso Plattner Institute of Design at Stanford [d.school], 2018; Stanford d.school, n.d.). Recent evidence suggests that DT-informed interventions can support engagement, problem solving, and student learning outcomes (Yu et al., 2024). Yet the institutionalization of such innovations is far more difficult than their initial design. DT can help generate promising curriculum, pedagogy, and assessment models, but it does not by itself specify the governance, QA, workload protection, moderation routines, or evidence systems required for institution-wide scale-up (Alzahmi et al., 2025; Mok & Han, 2017; Krooi et al., 2024).

A systematic review of graduate readiness and empirical research in Chinese HEIs situate meta-skills development within academic management (Che et al., 2026a, 2026e). Related research extends this work through AI literacy for academic management (Che, 2026b), institutional culture and student readiness in AI-enabled assessment (Che, 2026a), and integrated redesign of curriculum, pedagogy, and assessment (Che, 2026c). Complementary studies connect needs-based prioritization with implementation planning (Che et al., 2026b), translate diagnosed gaps into reform modules (Che et al., 2026d), and specify execution arrangements and governance for institutional transformation (Che et al., 2026c). These contributions provide a conceptual foundation for the present study's Delphi-validated implementation system linking reform design with institutional routines.

The present study addresses this “last-mile” problem. Building on prior framework-development work in Sichuan private HEIs (Che, 2024; Che, 2025), it proposes a governance and implementation playbook designed to translate design-oriented innovation into an institution-level operating logic. In practical terms, the playbook converts three upstream modules: agile curriculum integration, immersive experiential pedagogy, and smart meta-skills assessment: into a coordinated implementation system built around governance and decision rights, staged rollout, QA embedding, evidence routines and KPI indicators, and risk controls. Rather than introducing another conceptual model of meta-skills, the study focuses on how meta-skills reform can be organized, monitored, and sustained under real institutional constraints.

Figure 1 summarizes this logic. Conceptually, it links design thinking-derived innovation on the left to measurable meta-skills outcomes on the right through an institutional implementation layer in the center. The figure therefore contributes more than a visual summary: it makes explicit the study's core argument that innovation design is necessary but insufficient, and that sustainable meta-skills reform depends on complementary governance and evidence mechanisms that support institutionalization, comparability, and continuous improvement.

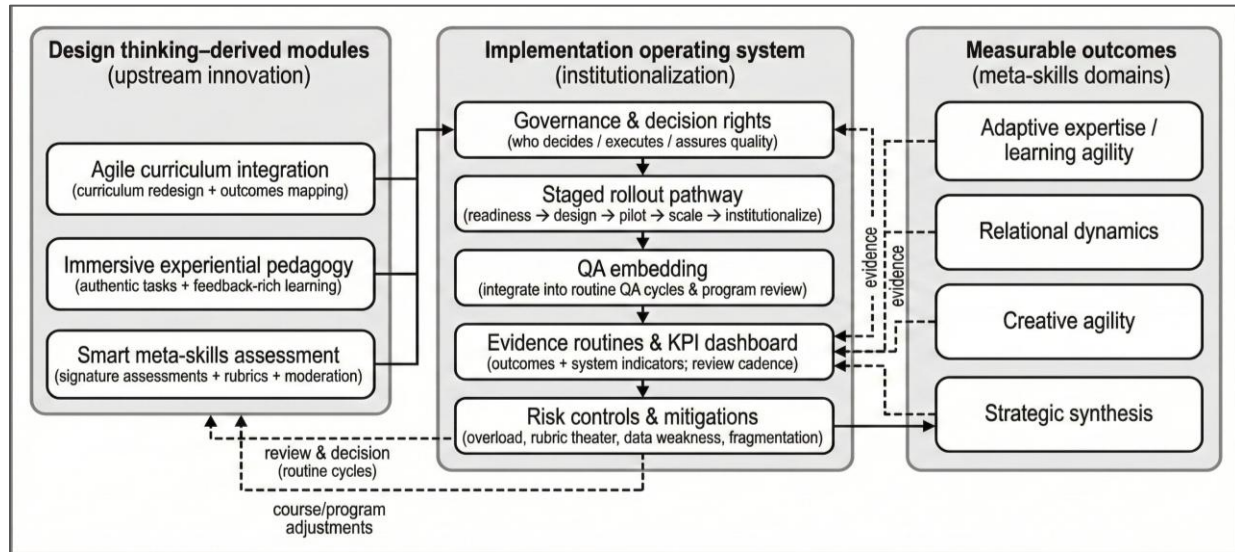


Figure 1 DT-derived modules to governance and evidence loops for meta-skills implementation

Guided by this framework, the study addresses three research questions:

RQ1: What governance architecture and decision-rights model best support institution-wide meta-skills implementation in Chinese private HEIs?

RQ2: What rollout stages, QA integration points, and KPI indicators achieve expert consensus for feasibility and sustainability?

RQ3: What implementation failure modes are judged most likely, and which mitigations are considered most actionable?

Together, these questions position the study's contribution as a validated implementation system for institutionalizing meta-skills, rather than as a further restatement of meta-skills as desirable educational aspiration.

Methodology

Research design

This study employed a two-round modified Delphi mixed-methods validation design to develop and refine a governance and implementation playbook for meta-skills-oriented academic management in Chinese private higher education institutions (HEIs), with an empirical focus on Sichuan Province. The design was sequential and implementation-oriented. First, a Version 1 playbook was developed by translating three design thinking-derived modules: agile curriculum integration, immersive experiential pedagogy, and smart meta-skills assessment: into institution-level operating mechanisms,

namely governance and decision rights, a staged rollout pathway, quality assurance (QA) embedding, KPI dashboard routines, and implementation risk controls. Second, the playbook was subjected to expert Delphi appraisal to assess whether each component was suitable for academic management practice in Sichuan private HEIs, feasible under resource- and workload-sensitive conditions, and sufficiently clear for execution. Quantitative consensus evidence informed component retention and revision, while qualitative feedback clarified sources of disagreement and guided targeted refinements, resulting in a finalized Version 2 playbook intended for institutional application. Because the expert panel drew on implementation experience across private HEIs in Sichuan Province, the findings should be interpreted as context-grounded rather than statistically generalizable; however, the resulting playbook is intended for analytic transfer to Chinese private HEIs operating under comparable governance and resource conditions.

Participants and instrument

A purposive Delphi panel of 15 experts was recruited to ensure coverage of the institutional functions required for system-wide implementation in Sichuan private HEIs. This panel size was considered appropriate because Delphi studies prioritize the relevance, heterogeneity, and information richness of expert judgment over statistical representativeness, particularly in operationally specialized policy and governance contexts (Hsu & Sandford, 2007; Jorm, 2015; Beiderbeck et al., 2021). Panel inclusion criteria were: (a) at least five years of professional experience in Chinese higher education leadership, academic governance, or quality assurance within Sichuan Province's private HEI sector; (b) substantive involvement in at least one implementation-relevant domain—curriculum redesign, assessment governance and moderation, QA and program review, teaching-learning enhancement, or institutional analytics and evidence use; and (c) demonstrated familiarity with implementation constraints typical of Sichuan private HEIs, including faculty workload pressures, limited implementation slack, compliance-oriented QA routines, and uneven data capacity across programs. Collectively, panel members reported direct professional involvement across eight private HEIs in Sichuan Province, providing breadth of institutional exposure while remaining consistent with Delphi logic, which emphasizes expert knowledge integration rather than population representativeness. The panel represented both strategic and operational perspectives, with a mean professional experience of 11.8 years (range = 6-22 years). In functional terms, panelists brought experience spanning institutional governance, program leadership, quality assurance, assessment oversight, teaching-learning enhancement, and educational analytics. To protect institutional confidentiality, participating HEIs are reported in aggregate rather than by name.

Data were collected using a structured Delphi instrument aligned with five playbook domains: (1) governance and decision rights, (2) staged rollout pathway, (3) QA embedding points, (4) evidence routines and KPI dashboard indicators, and (5) risk register with mitigations. Each component was rated on 5-point Likert scales for suitability, feasibility, and clarity (1 = very low; 5 = very high). KPI indicators and risk items were additionally ranked to identify early-stage implementation essentials. To strengthen content-validity reporting, the instrument also included relevance judgments enabling computation of item-level and scale-level content validity indices.

Procedures and analysis

The modified Delphi protocol followed a two-round iterative validation process to establish stability and convergence of expert judgments. In Round 1, panel members independently reviewed Version 1 and completed the rating and comment instrument, providing open-ended feedback on ambiguity, feasibility barriers, missing controls, and improvement suggestions relevant to implementation realities in Sichuan private HEIs. Between rounds, panelists received an anonymized summary of Round 1 domain-level statistics and synthesized qualitative themes to support reflection and convergence without direct panel interaction. Panelists did not receive item-level peer ratings; instead, they reviewed aggregated domain-level feedback and thematic summaries, which constitutes the principal modification from a more classical Delphi format (Keeney et al., 2001; Niederberger & Spranger, 2020; Spranger et al., 2022). Revisions were then documented through a structured change log linking each modification to quantitative dispersion patterns and qualitative rationales, producing Version 2. In Round 2, experts re-rated revised components and completed priority-ranking tasks for KPI indicators and implementation risks, yielding convergent evidence on feasibility, sustainability, and rollout sequencing.

Quantitative analysis was conducted at both item and domain levels. For each component, the study computed mean and standard deviation, median and interquartile range (IQR), and an agreement rate defined as the percentage of experts rating the component as 4 or 5. A component was classified as consensus-supported when it met at least two of three pre-specified criteria: median ≥ 4.00 , IQR ≤ 1.00 , and agreement $\geq 80\%$. Content validity evidence was reported using I-CVI, defined as the proportion of experts rating an item as relevant, and S-CVI/Ave, defined as the average I-CVI across items within a domain, using acceptability thresholds of I-CVI $\geq .80$ and S-CVI/Ave $\geq .90$. For ranked KPI and risk priorities, concordance was quantified using Kendall's coefficient of concordance (W), with W $\geq .70$ interpreted as strong agreement for managerial prioritization. Qualitative feedback was analyzed thematically using a structured coding framework focused on governance ambiguity and

decision rights, workload and capacity constraints, QA integration feasibility, assessment standardization and moderation, data integrity for dashboarding, and completeness of risk controls. Because qualitative coding was conducted by a single researcher for revision-oriented synthesis rather than interpretive theory generation, formal inter-coder reliability was not calculated; this should be regarded as a methodological limitation. Mixed-methods integration was implemented through a Delphi decision matrix pairing quantitative classifications (retain versus revise) with qualitative explanations and revision actions, ensuring that the finalized playbook was supported by both statistical convergence and context-grounded implementation reasoning applicable to private HEIs in Sichuan Province, China. All experts provided informed consent, and responses were anonymized and reported only in aggregate.

Findings

Overall consensus, content validity, and convergence across Delphi rounds

Across the two Delphi rounds, the expert panel showed clear convergence on the proposed playbook. Initial ratings were already high for suitability and feasibility, suggesting broad support for the playbook's managerial logic. However, Round 1 dispersion was greater for domains requiring cross-unit coordination and integration into routine institutional systems, particularly governance decision rights and QA embedding. These areas appear to have been more sensitive to familiar implementation constraints in the participating institutions, including limited coordination time, uneven moderation practices, and compliance-oriented QA routines. After revision, Round 2 ratings increased across all five domains, with narrower interquartile ranges and higher agreement rates. Content validity also remained strong: all domain-level S-CVI/Ave values exceeded the .90 acceptability threshold, and the proportion of items meeting I-CVI $\geq$.80 increased after revision (Table 1). Taken together, the Round 2 pattern indicates that Version 2 was judged not only suitable in principle but also more executable in practice, although feasibility remained slightly lower for QA embedding and evidence routines than for the other domains. Figure 2 illustrates this pattern.

Table 1 Domain-level consensus and content validity by playbook domain (Rounds 1-2; N = 15).

Playbook domain	Round	Suitability (M $\pm$ SD)	Feasibility (M $\pm$ SD)	Clarity (M $\pm$ SD)	Median	IQR	Agreement % (4–5)	S-CVI/Ave	I-CVI range
Governance & decision rights	R1	4.34 $\pm$ 0.48	4.08 $\pm$ 0.62	3.96 $\pm$ 0.66	4.0	1.0	78%	0.93	0.80–1.00
	R2	4.62 $\pm$ 0.39	4.40 $\pm$ 0.44	4.36 $\pm$ 0.50	5.0	1.0	89%	0.97	0.87–1.00
Staged rollout pathway	R1	4.49 $\pm$ 0.44	4.20 $\pm$ 0.53	4.11 $\pm$ 0.57	4.0	1.0	84%	0.95	0.80–1.00
	R2	4.70 $\pm$ 0.32	4.51 $\pm$ 0.38	4.44 $\pm$ 0.44	5.0	1.0	92%	0.98	0.87–1.00

Playbook domain	Round	Suitability (M ± SD)	Feasibility (M ± SD)	Clarity (M ± SD)	Median	IQR	Agreement % (4–5)	S-CVI/Ave	I-CVI range
QA embedding	R1	4.29 ± 0.51	3.98 ± 0.66	3.90 ± 0.69	4.0	1.0	76%	0.92	0.80–0.93
	R2	4.55 ± 0.40	4.28 ± 0.48	4.31 ± 0.49	4.0	1.0	88%	0.96	0.87–1.00
Evidence routines & KPI dashboard	R1	4.41 ± 0.46	4.06 ± 0.61	4.05 ± 0.60	4.0	1.0	81%	0.94	0.80–1.00
	R2	4.63 ± 0.35	4.36 ± 0.42	4.39 ± 0.45	5.0	1.0	90%	0.97	0.87–1.00
Risk register & mitigations	R1	4.38 ± 0.49	4.19 ± 0.55	4.07 ± 0.60	4.0	1.0	83%	0.94	0.80–1.00
	R2	4.59 ± 0.37	4.41 ± 0.41	4.33 ± 0.46	5.0	1.0	91%	0.97	0.87–1.00

Note. Consensus-supported indicates that a domain met at least two of three pre-specified criteria: median ≥ 4.00 , IQR ≤ 1.00 , and agreement $\geq 80\%$. Agreement refers to the percentage of experts assigning ratings of 4 or 5.

Revisions mainly targeted items that had narrowly missed the agreement threshold in Round 1 or showed lower clarity. By Round 2, all domains met the study’s consensus criteria, with medians of at least 4.0, IQRs of 1.0 or below, and agreement rates ranging from 88% to 92%. This pattern suggests that the revisions improved operational clarity and brought previously borderline items into an acceptable range of expert support.

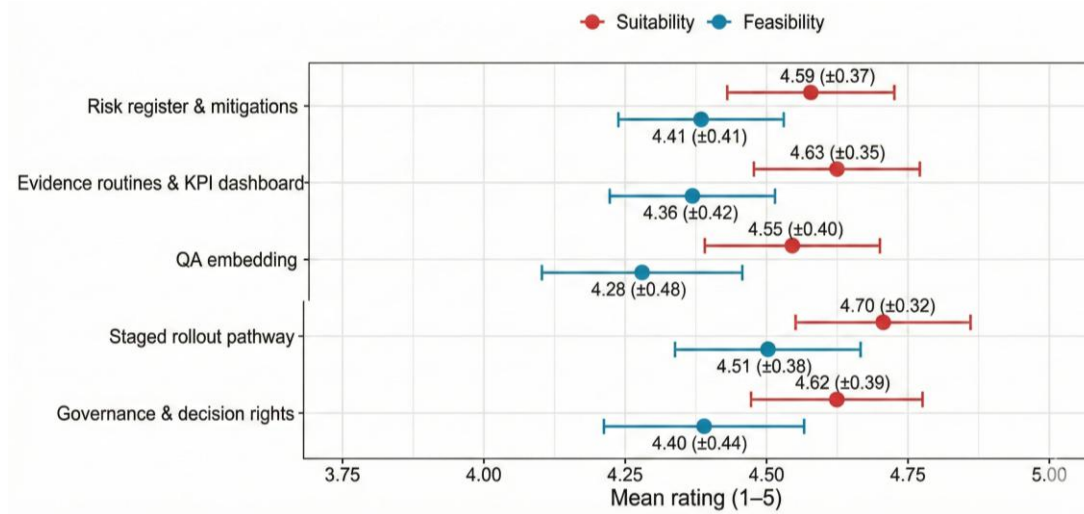


Figure 2 Round 2 ratings of suitability and feasibility by playbook domain

Qualitative feedback pointed to four recurring refinement themes: clearer decision rights, workload protection and role rotation, minimum standards for moderation and evidence capture, and integration of meta-skills evidence into existing QA artifacts rather than parallel systems. These themes help explain why feasibility improved after revision. They also indicate that, in the target context, implementation depends less on adding new structures than on reducing coordination burden while

strengthening evidence reliability.

Governance architecture and decision rights for institutional scale (RQ1)

Round 2 findings suggest that governance clarity is central to implementation stability. All six governance components met the consensus-supported classification, indicating broad agreement that institution-wide meta-skills implementation requires formal coordination structures rather than reliance on isolated champions. The strongest support was observed for the steering committee's role in standard setting, moderation policy, and cross-program calibration. This pattern suggests that experts viewed comparability as a foundational condition for implementation: without shared standards and routine calibration, meta-skills outcomes risk becoming program-specific and therefore less auditable and less improvable.

Round 1 comments also identified two governance tensions: balancing program autonomy with institution-level standardization, and managing conflicts over rubrics, workload, and resource priorities. Version 2 addressed these issues by making decision rights more explicit. Program design cells were positioned as responsible for local mapping and assessment drafting, the steering committee for standards and cross-program comparability, and the executive sponsor group for resourcing and dispute resolution. As shown in Table 2, all governance components were retained, with especially strong support for mechanisms linked to comparability, monitoring, and conflict resolution.

Table 2 Governance and decision-rights components: Round 2 consensus and decision classification (N = 15; Sichuan private HEIs).

Governance component	Suitability (M ± SD)	Feasibility (M ± SD)	Clarity (M ± SD)	Median	IQR	Agreement % (4–5)	Decision
Executive sponsor group (resourcing, conflict resolution)	4.60 ± 0.49	4.33 ± 0.56	4.27 ± 0.57	5.0	1.0	87%	Retain
Meta-skills steering committee (standards, monitoring, alignment)	4.73 ± 0.44	4.53 ± 0.51	4.47 ± 0.52	5.0	1.0	93%	Retain
Program design cells (mapping, signature assessments, workload plan)	4.58 ± 0.51	4.40 ± 0.51	4.38 ± 0.49	4.0	1.0	89%	Retain
Teaching innovation support (PD, coaching, templates)	4.53 ± 0.52	4.33 ± 0.56	4.32 ± 0.55	4.0	1.0	88%	Retain
Evidence & analytics function (dashboard, audits, integrity checks)	4.67 ± 0.49	4.40 ± 0.51	4.36 ± 0.50	5.0	1.0	91%	Retain
Escalation pathway (rubric disputes, workload conflicts)	4.53 ± 0.52	4.20 ± 0.68	4.27 ± 0.61	4.0	1.0	84%	Retain

Overall, the governance results support the argument that sustainable implementation depends on explicit decision rights and routine cross-program coordination. In particular, the data indicate that assessment comparability and evidence integrity are treated by experts not as secondary QA concerns, but as core implementation conditions.

Rollout pathway and QA embedding (RQ2)

The panel strongly endorsed a staged implementation pathway with explicit go/no-go criteria, especially in the readiness and pilot phases. In Round 2, staged rollout showed the highest feasibility rating among the five domains ($M = 4.51$, $SD = 0.38$; Table 1), suggesting that sequencing was seen as a practical way to reduce overload and allow calibration before expansion. Experts also emphasized that pilot implementation should include moderation and artifact sampling requirements so that rubric use can be stabilized before scaling.

Feasibility for QA embedding improved from 3.98 in Round 1 to 4.28 in Round 2, largely because Version 2 positioned QA as integration into existing review structures rather than as an added documentation layer. The panel also endorsed a balanced KPI dashboard that combined outcome indicators with system indicators, indicating that experts viewed implementation monitoring as requiring both evidence of student performance and evidence of process quality. Priority rankings showed strong concordance, suggesting that experts agreed not only on what was important, but also on what should be prioritized first in practice.

Table 3 KPI and implementation risk priorities (Round 2; $N = 15$; Sichuan private HEIs).

Priority set	Kendall's W	p-value	Highest-priority items (top 5) with mean priority (1–5)
KPI indicators	0.78	< .001	(1) Meta-skills rubric scores on signature assessments (4.73); (2) Moderation completion rate (4.60); (3) Curriculum mapping coverage (% courses aligned) (4.53); (4) Inter-rater agreement/calibration completion (4.47); (5) Equity/variance check by program (4.33)
Implementation risks	0.74	< .001	(1) Faculty overload/insufficient protected time (4.80); (2) Rubric theater (rubrics exist but not used for feedback) (4.60); (3) Data weakness/missing artifacts (4.53); (4) Fragmented implementation across programs (4.47); (5) Student resistance to authenticity/workload (4.20)

Note. Kendall's W values closer to 1.00 indicate stronger rank concordance. In this study, $W \geq .70$ was interpreted as strong agreement for managerial prioritization.

Qualitative feedback repeatedly linked feasibility to two practical controls: protected workload allocation and disciplined moderation routines. Experts treated these as minimum operating conditions for credible meta-skills evidence. Without them, dashboards would be less reliable, QA processes could become symbolic, and cross-program comparability would weaken. Figure 3 illustrates how readiness checks, pilot calibration, and routine review points are intended to support sustained scale-up.

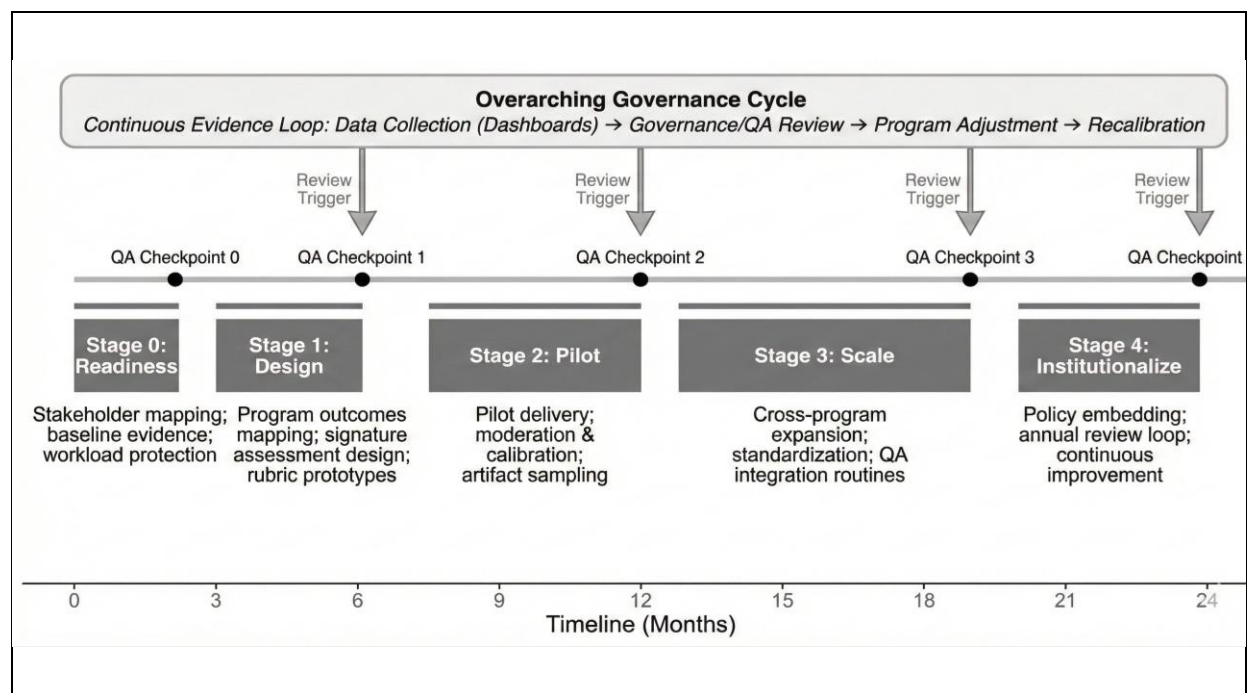


Figure 3 Staged rollout pathway with QA checkpoints and evidence loops for institutional scale-up

Overall, the findings show a consistent cross-cutting pattern: experts endorsed meta-skills implementation most strongly when the playbook reduced coordination ambiguity, embedded QA within existing routines, and protected the integrity of evidence used for review and improvement. RQ1 is supported by consensus on governance and decision rights, RQ2 by convergence on rollout sequencing, QA embedding, and KPI priorities, and RQ3 by a stable hierarchy of implementation risks and mitigations. Overall, the results suggest that scalable meta-skills implementation depends less on introducing new aspirations than on specifying the institutional conditions under which those aspirations can be governed, monitored, and sustained.

Implementation risks and mitigations (RQ3)

Risk priorities showed similarly strong agreement (Kendall's $W = 0.74$, $p < .001$), indicating that experts shared a stable view of the most likely implementation threats. The highest-ranked risks were faculty overload, rubric theater, weak data capture, fragmented implementation across programs, and student resistance to authenticity-related workload. These risks point to a common pattern: implementation failure was viewed less as a problem of conceptual disagreement than as a problem of insufficient protected time, weak moderation discipline, and uneven evidence routines. The endorsed mitigations: protected workload, shared exemplars, sample audits, and minimum evidence

standards: therefore function as practical safeguards against symbolic compliance and implementation drift.

Discussion

This study addressed a persistent last-mile problem in meta-skills reform in Chinese private HEIs: innovation models may generate promising curriculum, pedagogy, and assessment ideas, but these rarely become sustainable unless they are supported by explicit governance, quality assurance, and evidence routines. The Delphi results suggest that the proposed playbook was judged not only desirable in principle but workable under the resource and workload conditions typical of the target context. In this sense, the findings shift meta-skills from aspirational graduate attributes to governable outcomes that require institutional standards, comparable evidence, and routine review.

For RQ1, the strongest consensus concerned governance and decision rights. Experts placed particular importance on the steering committee's role in standard setting and calibration, as well as the evidence or analytics function's role in integrity checking. This pattern suggests that cross-program comparability is a foundational condition of institutionalization. Without formal decision rights and routine calibration, programs are likely to drift toward local interpretations of meta-skills and rubric use, weakening both QA credibility and the usefulness of dashboard evidence. The clearer division in Version 2 between program-level design autonomy and institution-level standards therefore appears to resolve a central tension in multi-program private HEIs.

For RQ2, staged rollout was endorsed as a feasibility safeguard rather than a delay mechanism. In the target context, sequencing appears valuable because it reduces overload, creates time for calibration, and lowers the risk of fragmented uptake. The improvement in QA embedding after revision is especially revealing. Experts did not resist QA integration itself; they resisted additional documentation burden. Once QA was reframed as integration into existing artifacts rather than as a parallel reporting system, feasibility increased. The KPI pattern reinforces the same logic: credible outcome evidence depends on process integrity, including moderation completion, curriculum mapping coverage, artifact availability, and equity or variance monitoring. Meta-skills become improvable only when institutions manage both outcomes and the reliability of the evidence system that supports them.

For RQ3, the ranked risks: faculty overload, rubric theater, weak data capture, and cross-program fragmentation: indicate that implementation failure is likely to arise less from disagreement over the value of meta-skills than from weak operating conditions. The endorsed mitigations: protected

workload, role rotation, minimum evidence standards, moderation routines, and shared templates: therefore function as a practical control bundle against symbolic compliance and implementation drift. This is an important point: in the participating institutions, credible reform was judged to depend more on managerial discipline than on expensive technological expansion.

The study also clarifies the role of design thinking in institutional reform. The findings support design thinking as a useful upstream method for generating innovation, but not as a sufficient model for downstream institutionalization. The contribution of this paper is therefore not the invention of entirely new mechanisms. Many components of the playbook: staged rollout, moderation routines, and KPI monitoring: are already familiar within QA practice. The novelty lies in assembling them into a validated implementation logic explicitly oriented toward meta-skills as institutional outcomes in Chinese private HEIs. The contribution is thus primarily contextual and integrative rather than radically theoretical.

The practical implications are straightforward. For institutional leaders, readiness for adoption should be judged not by policy enthusiasm alone but by the presence of clear decision rights, protected workload, moderation capacity, and minimum evidence routines. For QA leaders and policymakers, the findings suggest that meta-skills reform becomes more credible when aligned with routine review structures and when dashboarding combines outcome evidence with process-integrity indicators. Although the study was conducted from an institutional management perspective, its ultimate significance is student-facing: stronger moderation, clearer assessment expectations, and more reliable evidence routines are likely to support more consistent learning experiences and more trustworthy feedback across programs.

At the same time, the findings require cautious interpretation. As in many modified Delphi studies, the upward movement from Round 1 to Round 2 may reflect not only improved clarity but also anchoring or social desirability effects after structured feedback. The study therefore provides validation evidence on feasibility, priorities, and risk controls rather than direct evidence of implementation effectiveness or student outcome gains. Future research should test the playbook in live multi-site settings, examine implementation fidelity and evidence reliability over time, and assess whether stronger governance and moderation routines are associated with measurable gains in student meta-skills.

Finally, the transferability boundary should be stated narrowly. The playbook is expected to apply most directly to Chinese private HEIs with multi-program structures, moderate QA maturity, limited protected time for redesign uneven moderation capacity, and pressure to demonstrate

employability-oriented learning outcomes under constrained resources. Within such settings, the study offers a defensible and practically useful logic for moving from innovation prototypes to institutionally sustained meta-skills implementation.

Conclusion

This study addressed a persistent last-mile problem in Chinese private higher education: meta-skills are widely endorsed as important graduate outcomes, yet they often remain aspirational because curriculum, pedagogy, assessment, and quality assurance are not coordinated through a workable institutional system. In response, the study developed and validated a practical implementation playbook that translates meta-skills goals into explicit governance routines, staged rollout, QA integration, measurable indicators, and risk controls suited to the resource and workload conditions typical of Sichuan private HEIs. The two-round modified Delphi results showed strong expert support for the revised playbook, with improved clarity after revision, high perceived suitability and feasibility across domains, and strong content validity. The findings further indicate that early implementation depends on building credible evidence and comparability through signature assessments, moderation and calibration routines, curriculum mapping coverage, and equity or variance monitoring, while the most significant threats arise from practical management failures such as faculty overload, rubric theater, weak data capture, and fragmented cross-program implementation. In sum, the study contributes not a new meta-skills framework, but a validated operating logic for institutionalizing meta-skills as governable outcomes. The playbook is expected to be most useful in Chinese private HEIs with comparable QA maturity, moderation capacity, and resource constraints. Future research should test the playbook through live multi-site implementation, examine implementation fidelity and evidence reliability over time, and determine whether stronger governance and QA routines are associated with measurable gains in student meta-skills development.

Funding

The author(s) declared that financial support was received for this work and/or its publication. The work was conducted as a self-funded initiative with administrative approval from Payap University (International College, Educational Leadership Program), Project No. 04-08-03-68-003-01.

References

- Alzahmi, R. A., Syed, R. T., Singh, D., Arshi, T. A., & Kutty, S. V. (2025). Organizational change in higher education institutions: Thematic mapping of the literature and future research agenda. *Humanities & Social Sciences Communications*, 12, Article 1282. <https://doi.org/10.1057/s41599-025-05650-w>
- Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, 37, Article 100489. <https://doi.org/10.1016/j.edurev.2022.100489>
- Beiderbeck, D., Frevel, N., von der Gracht, H. A., Schmidt, S. L., & Schweitzer, V. M. (2021). Preparing, conducting, and analyzing Delphi surveys: Cross-disciplinary practices, new directions, and advancements. *MethodsX*, 8, Article 101401. <https://doi.org/10.1016/j.mex.2021.101401>
- Brown, P., Sadik, S., & Xu, J. (2021). Higher education, graduate talent and the prospects for social mobility in China's innovation nation. *International Journal of Educational Research*, 109, Article 101841. <https://doi.org/10.1016/j.ijer.2021.101841>
- Che, C. (2024). *Academic management innovation based on the concept of meta skills for private higher education institutions in Sichuan, China* [Doctoral dissertation, Chulalongkorn University]. <https://doi.org/10.58837/CHULA.THE.2024.486>
- Che, C. (2025). *Next-gen academic management: A strategic framework for meta-skills integration in Sichuan private HEIs—Empowering 21st-century competency and institutional innovation* [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.17827034>
- Che, C. (2026a). AI-enabled assessment in higher education: Institutional culture and student meta-skill readiness across university tiers. *Social Sciences & Humanities Open*, 14, Article 103141. <https://doi.org/10.1016/j.ssaho.2026.103141>
- Che, C. (2026b). AI literacy as a meta-skill: A four-domain model for academic management innovation in higher education. *Frontiers in Education*, 11, Article 1755238. <https://doi.org/10.3389/feduc.2026.1755238>
- Che, C. (2026c). Institutional design thinking: Re-architecting curriculum, pedagogy and assessment for meta-skills in Chinese private HEIs. *International Journal of Learning, Teaching and Educational Research*, 25(5), 1-25. <https://doi.org/10.26803/ijlter.25.5.1>

- Che, C., Chaemchoy, S., & Siribanpitak, P. (2026a). Academic management and meta-skills in Chinese higher education: A systematic review of graduate readiness. *Journal of Education and Learning (EduLearn)*, 20(1), 264-277. <https://doi.org/10.11591/edulearn.v20i1.23883>
- Che, C., Chaemchoy, S., & Siribanpitak, P. (2026b). Meta-skills-oriented academic management in higher education: Evidence from Chinese HEIs. *International Journal of Evaluation and Research in Education*, 15(1), 327-341. <https://doi.org/10.11591/ijere.v15i1.37685>
- Che, C., Chaemchoy, S., & Siribanpitak, P. (2026c). Empathy-to-Execution (E2E): An execution-oriented design thinking model for meta-skills-based academic system transformation. *Journal of Education and Innovation*, 28(2), Article 290897. <https://doi.org/10.71185/jeiejournals.v28i1.290897>
- Che, C., Chaemchoy, S., & Siribanpitak, P. (2026d). From gaps to modules: Design thinking for meta-skills in Chinese private universities. *Journal of Education and Learning (EduLearn)*, 20(3), 1803-1816. <https://doi.org/10.11591/edulearn.v20i3.26930>
- Che, C., Chaemchoy, S., & Siribanpitak, P. (2026e). Meta-skills-oriented academic management in higher education: Evidence from Chinese HEIs. *International Journal of Evaluation and Research in Education*, 15(1), 327-341. <https://doi.org/10.11591/ijere.v15i1.37685>
- Chen, L., Geng, X., Lu, M., Shimada, A., & Yamada, M. (2023). How students use learning analytics dashboards in higher education: A learning performance perspective. *SAGE Open*, 13(3). <https://doi.org/10.1177/21582440231192151>
- du Plooy, E., Casteleijn, D., & Franzsen, D. (2024). Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement. *Heliyon*, 10(21), Article e39630. <https://doi.org/10.1016/j.heliyon.2024.e39630>
- Elkington, S., & Chesterton, P. (2025). Embedding assessment flexibilities for future authentic learning. *Teaching in Higher Education*, 30(3), 700-716. <https://doi.org/10.1080/13562517.2024.2394605>
- Gravett, K. (2025). Authentic assessment as relational pedagogy. *Teaching in Higher Education*, 30(3), 608-622. <https://doi.org/10.1080/13562517.2024.2380997>
- Han, W. (2025). Understanding creative teaching in China: An empirical study by a narrative inquiry method. *Thinking Skills and Creativity*, 56, Article 101761. <https://doi.org/10.1016/j.tsc.2025.101761>
- Harris, A., Jones, M., & Ismail, N. (2022). Distributed leadership: Taking a retrospective and contemporary view of the evidence base. *School Leadership & Management*, 42(5), 438-456. <https://doi.org/10.1080/13632434.2022.2109620>

- Hasso Plattner Institute of Design at Stanford (d.school). (2018). *Design thinking bootleg*. Stanford University. <https://dschool.stanford.edu/resources/design-thinking-bootleg>
- Hsu, C.-C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research, and Evaluation*, 12(1), Article 10. <https://doi.org/10.7275/pdz9-th90>
- Janse van Rensburg, C., Coetzee, S. A., & Schmulian, A. (2022). Developing digital creativity through authentic assessment. *Assessment & Evaluation in Higher Education*, 47(6), 857-877. <https://doi.org/10.1080/02602938.2021.1968791>
- Jorm, A. F. (2015). Using the Delphi expert consensus method in mental health research. *Australian & New Zealand Journal of Psychiatry*, 49(10), 887-897. <https://doi.org/10.1177/0004867415600891>
- Keeney, S., Hasson, F., & McKenna, H. P. (2001). A critical review of the Delphi technique as a research methodology for nursing. *International Journal of Nursing Studies*, 38(2), 195-200. [https://doi.org/10.1016/S0020-7489\(00\)00044-4](https://doi.org/10.1016/S0020-7489(00)00044-4)
- Krooi, M., Whittingham, J., & Beausaert, S. (2024). Introducing the 3P conceptual model of internal quality assurance in higher education: A systematic literature review. *Studies in Educational Evaluation*, 82, Article 101360. <https://doi.org/10.1016/j.stueduc.2024.101360>
- McArthur, J. (2023). Rethinking authentic assessment: Work, well-being, and society. *Higher Education*, 85, 85-101. <https://doi.org/10.1007/s10734-022-00822-y>
- Mok, K. H., & Han, X. (2017). Higher education governance and policy in China: Managing decentralization and transnationalism. *Policy and Society*, 36(1), 34-48. <https://doi.org/10.1080/14494035.2017.1288964>
- Nadeem, M. (2024). Distributed leadership in educational contexts: A catalyst for school improvement. *Social Sciences & Humanities Open*, 9, Article 100835. <https://doi.org/10.1016/j.ssaho.2024.100835>
- Niederberger, M., & Spranger, J. (2020). Delphi technique in health sciences: A map. *Frontiers in Public Health*, 8, Article 457. <https://doi.org/10.3389/fpubh.2020.00457>
- Nieminen, J. H., Bearman, M., & Ajjawi, R. (2023). Designing the digital in authentic assessment: Is it fit for purpose? *Assessment & Evaluation in Higher Education*, 48, 529-543. <https://doi.org/10.1080/02602938.2022.2089627>
- Nieminen, J. H., Haataja, E., & Cobb, P. J. (2024). From active learners to knowledge contributors: Authentic assessment as a catalyst for students' epistemic agency. *Teaching in Higher Education*, 30, 970-990. <https://doi.org/10.1080/13562517.2024.2332252>

- Otermans, P., Small, S., Pereira, M., Sridharan, P., & Aditya, D. (2025). Students' and graduates' perspectives on skill development in higher education. *Journal of Education and Work*, 38, 84-112. <https://doi.org/10.1080/13639080.2025.2576858>
- Prasittichok, P., & Klaykaew, K. K. (2022). Meta-skills development needs assessment among undergraduate students. *Heliyon*, 8(1), Article e08787. <https://doi.org/10.1016/j.heliyon.2022.e08787>
- Sokhanvar, Z., Salehi, K., & Sokhanvar, F. (2021). Advantages of authentic assessment for improving the learning experience and employability skills of higher education students: A systematic literature review. *Studies in Educational Evaluation*, 70, Article 101030. <https://doi.org/10.1016/j.stueduc.2021.101030>
- Sorour, A., & Atkins, A. S. (2024). Big data challenge for monitoring quality in higher education institutions using business intelligence dashboards. *Journal of Electronic Science and Technology*, 22(1), Article 100233. <https://doi.org/10.1016/j.jnlest.2024.100233>
- Spranger, J., Homberg, A., Sonnberger, M., & Niederberger, M. (2022). Reporting guidelines for Delphi techniques in health sciences: A methodological review. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 172, 1-11. <https://doi.org/10.1016/j.zefq.2022.04.025>
- Stanford d.school. (n.d.). *An introduction to design thinking: Process guide* [PDF]. Stanford University. <https://web.stanford.edu/~mshanks/MichaelShanks/files/509554.pdf>
- Tuononen, T., Hyytinen, H., Kleemola, K., Hailikari, T., Männikkö, I., & Toom, A. (2022). Systematic review of learning generic skills in higher education—Enhancing and impeding factors. *Frontiers in Education*, 7, Article 885917. <https://doi.org/10.3389/feduc.2022.885917>
- Villarroel, V., Melipillán, R., Santana, J., & Aguirre, D. (2024). How authentic are assessments in vocational education? An analysis from Chilean teachers, students and examinations. *Frontiers in Education*, 9, Article 1308688. <https://doi.org/10.3389/feduc.2024.1308688>
- Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in Educational Evaluation*, 83, Article 101425. <https://doi.org/10.1016/j.stueduc.2024.101425>
- Xu, B. (2021). Policy and governance of higher education in China. *ECNU Review of Education*. <https://doi.org/10.1177/22125868211046032>

Yu, Q., Yu, K., & Lin, R. (2024). A meta-analysis of the effects of design thinking on student learning. *Humanities & Social Sciences Communications*, 11, Article 742. <https://doi.org/10.1057/s41599-024-03237-5>

Zhan, Y., Boud, D., & Du, Z. (2025). Designing for authentic assessment: A scoping review. *Higher Education*. Advance online publication. <https://doi.org/10.1007/s10734-025-01588-9>