iEducation, 2026, 2(3), 47-61
https://doi.org/10.71204/3n87y230

From Apprenticeship to Accountable Skill: A Simulation-Based Curriculum Framework for Neurosurgical Training in China

Haoxiang Xu [1,†], Zirui Zhang[2, †], Chengyin Lu [1], Chao Ma [1], Xinjie Hong [1,*], Danfeng Zhang [1,*]

[1] Department of Neurosurgery, Shanghai Changzheng Hospital, Naval Medical University, Shanghai 200003, China

[2] Department of Clinical Medicine, Naval Medical University, Shanghai 200003, China

† Haoxiang Xu and Zirui Zhang contributed equally to this work.

*** Correspondence:**

Xinjie Hong

cnhxjie@126.com

Danfeng Zhang

dfzhangsmmu@163.com

Received: 1 October 2026 /Accepted: 11 November 2026 / Published online: 12 November 2026

Abstract

Simulation offers neurosurgical training a way to rehearse high-risk procedures without transferring the earliest part of the learning curve to patients. The central educational problem is therefore no longer simply whether simulators work; it is how to organize them into a coherent curriculum framework that connects practice with readiness for supervised operative responsibility. This article analyzes that organizational challenge for China. International experience identifies three conditions for effective simulation-based education: procedure-level competency expectations, models selected for instructional value rather than fidelity alone, and assessment systems whose scores are supported by validity evidence. Published evidence from China documents a standardized neurosurgical training pathway with a dedicated microsurgical-skills rotation, but it does not establish how simulation is distributed, taught, or used to inform progression. We therefore outline a curriculum framework organized around four functions: defining entrustable technical activities, matching infrastructure to educational yield, developing a distributed instructor and rater network, and embedding digital learning within a shared evaluation system. The argument shifts the discussion from acquiring equipment to building an accountable training system. A staged implementation process-beginning with shared standards and pilot centers, expanding through regional instructor networks, and maturing through registry-linked evaluation-could make simulation a practical mechanism for safer and more equitable neurosurgical training in China.

Keywords: Neurosurgical Education; Surgical Simulation; Entrustment; Curriculum Development

1. Introduction

Neurosurgery is among the most technically demanding medical specialties. Operations on the brain and spinal cord offer narrow margins for error, prolonged learning curves, and limited tolerance for the trial-and-error learning that historically characterized surgical apprenticeship. At the same time, the environment in which neurosurgeons are trained has changed substantially. Duty-hour restrictions, heightened patient-safety expectations, and the increasing technical complexity of procedures have reduced the operative autonomy that trainees can safely be granted, prompting educators worldwide to seek complementary training methods that move part of the learning curve outside the operating room (Ganju et al., 2013; Selden et al., 2012).

Simulation-based education has emerged as the most prominent of these complements. In its broadest sense, simulation encompasses synthetic bench models, human and animal tissue, three-dimensionally printed replicas, and computer-based virtual reality (VR) systems with or without haptic feedback (Davids et al., 2021; Delorme et al., 2012). Evidence from general surgery first demonstrated that skills acquired on VR simulators could transfer to the operating room (Seymour et al., 2002), and subsequent systematic reviews confirmed measurable skills transfer across surgical disciplines (Sturm et al., 2008). Within neurosurgery specifically, systematic reviews now document a rapidly expanding portfolio of simulators and generally favorable effects on procedural knowledge and technical performance (Davids et al., 2021; Patel et al., 2020). Parallel developments in educational theory—competency-based medical education (CBME), entrustable professional activities (EPAs), and milestone frameworks—have provided the conceptual scaffolding needed to integrate simulation into formal curricula rather than treating it as an optional enrichment activity (Frank et al., 2010; ten Cate, 2005; ten Cate et al., 2015).

China gives this discussion a specific institutional setting. The national neurosurgical training pathway already includes a microsurgical-skills laboratory rotation, providing a formal point at which structured technical instruction can be anchored (Xu et al., 2016). What remains unclear is whether simulation elsewhere in the pathway follows shared curricular goals, whether trainees across regions receive comparable access to useful models and instructors, and whether simulator performance influences later workplace entrustment. In other words, the existence of a skills laboratory is not itself evidence that simulation functions as a curriculum.

Against this background, the present article asks a narrower organizational question: how can simulation be converted from scattered equipment and local enthusiasm into a dependable component of neurosurgical training across Chinese institutions? The discussion begins with what is known—and not known—about Chinese practice, then considers international experience in competency design, simulator selection, and measurement. The final section translates those lessons into a curriculum framework suited to China's governance and institutional context. The emphasis throughout is on readiness for entrustment and educational accountability rather than on technology adoption as an end in itself. Where the supporting evidence is indirect or drawn from other specialties, that limitation is stated explicitly rather than converted into an unsupported claim of learner benefit.

2. Chinese Foundations and Evidence Gaps

China has moved toward a standardized national neurosurgical training structure, and the national residency curriculum incorporates a dedicated microsurgical-skills laboratory rotation (Xu et al., 2016). This foundation is an important asset: it means that the institutional concept of a skills laboratory is already codified in the national training structure, and that national bodies have established mechanisms for setting and enforcing training standards that could, in principle, be extended to simulation-based education.

In practice, however, published evidence is insufficient to determine how evenly simulation has penetrated everyday neurosurgical training across Chinese institutions. Whether tertiary centers in major cities operate substantially better-equipped simulation facilities than regional training bases, and how many training bases have dedicated simulation space, consumables, or protected teaching time, has not been systematically documented. The cost of high-fidelity equipment is a recognized barrier internationally—surveys of United States program directors identified expense and availability as the principal obstacles to broader simulation adoption (Ganju et al., 2013)—and how these obstacles weigh in a system that must train a very large neurosurgical workforce is itself an open empirical question. Experiences from other resource-conscious settings are instructive: a national-level, simulation-based neurosurgery bootcamp developed outside a high-income context documented both feasibility and detailed cost structures, illustrating that meaningful simulation programs need not depend on the most expensive platforms (Bakhshi et al., 2022).

A second challenge is curricular rather than material. Published evidence is insufficient to determine whether simulation use in Chinese neurosurgical programs is embedded in structured, criterion-referenced curricula with defined proficiency targets, or remains opportunistic and dependent on individual faculty initiative. Chinese medical schools have experimented with instructional reform in neurosurgery; one single-institution study reported that case-based and problem-based teaching reforms were associated with improved student engagement and recruitment into the specialty (Shen et al., 2022), but comparable published work describing simulation-integrated technical curricula, their assessment instruments, or their outcomes in Chinese neurosurgical training is scarce. Consequently, published evidence is insufficient to determine what proportion of Chinese neurosurgical residents achieve defined technical milestones before performing procedures on patients, or how technical competence is currently certified.

Third, published evidence is insufficient to determine the state of assessment infrastructure in Chinese neurosurgical training. Internationally, technical skills are increasingly measured with validated instruments rather than global subjective impressions, yet we found no published reports of Chinese neurosurgical programs routinely employing such tools, and no national repository of benchmark performance data appears to have been described. Similarly, published evidence is insufficient to determine the availability of faculty development for simulation teaching and rater training; reliable scoring of technical performance requires trained raters, and the inter-rater reliability of neurosurgical assessment scales cannot be assumed without deliberate calibration exercises.

Finally, the evaluation literature within China is thin. International commentators have repeatedly noted that much of the global simulation literature reports immediate, simulator-based outcomes rather than transfer to clinical performance or patient outcomes, and that reporting quality is variable (Cook et al., 2013; McGaghie et al., 2011). These limitations apply with even greater force where programs are new. Any Chinese expansion of simulation-based training should therefore be coupled from the outset to transparent evaluation and reporting, so that domestic evidence accumulates rather than relying indefinitely on extrapolation from North American and European studies.

3. What International Experience Can—and Cannot—Teach

Simulation-based neurosurgical education in North America and Europe has matured along three interacting lines: competency frameworks that define what training should achieve; simulation modalities that provide the practice environment; and assessment and instructional-design science that determines whether practice translates into competence. Each line offers transferable lessons—and caveats—for China.

3.1. Competency and Entrustment

Competency-based medical education reframes training around the outcomes a graduate must demonstrably achieve, rather than the time served in rotations (Frank et al., 2010). Within CBME, the concept of entrustable professional activities (EPAs) links abstract competencies to concrete units of clinical work that a supervisor can progressively entrust to a trainee as observed competence accrues (ten Cate, 2005). AMEE guidance has codified how EPAs can be constructed into workplace-based curricula, emphasizing explicit entrustment decisions and graded supervision (ten Cate et al., 2015).

In neurosurgery, the most visible implementation of this thinking is the Neurological Surgery Milestones, developed jointly by the Accreditation Council for Graduate Medical Education (ACGME) and the Society of Neurological Surgeons and subsequently revised (ACGME & Society of Neurological Surgeons, 2018). The milestones describe developmental trajectories across patient care and technical subcompetencies and are used in the United States as an official framework for semiannual resident assessment. Two cautions accompany their use. First, the milestones are a reporting and developmental framework, not a validated certification standard; they were not designed to function as high-stakes pass/fail criteria. Second, the framework is jurisdiction-specific: it reflects United States training structures and cannot simply be transplanted into another national system without adaptation.

Specialty-specific EPA work in neurosurgery remains limited but is growing. In vascular neurosurgery, for example, an international group defined a set of EPAs for neurovascular microsurgery training, covering activities from microanastomosis to aneurysm surgery with specified levels of entrustment (van Lieshout et al., 2023). Such work illustrates how global competency language can be operationalized into procedure-level expectations that simulation laboratories are well placed to support: low-stakes rehearsal, deliberate practice, and formative entrustment decisions can all occur in the simulated environment before graded clinical entrustment.

The key transferable insight for China is that simulation achieves its full value when it is anchored to an explicit competency framework; equipment without curriculum produces activity, not competence.

3.2. Choosing Simulation Modalities by Educational Value

The neurosurgical simulation portfolio now spans several broad categories. Physical bench and synthetic models support part-task practice such as microanastomosis, drilling, and endoscopic navigation. Biological models—including cadaveric specimens and, at the lower-cost end, models using human placenta for bypass and aneurysm-clipping practice—offer realistic tissue handling (Belykh et al., 2017). Three-dimensionally printed, patient-specific replicas permit rehearsal of actual operative anatomy; a patient-specific aneurysm clipping model produced by 3D printing and silicone casting is a representative example (Ryan et al., 2016). Virtual reality systems, from early physics-based cranial microneurosurgery simulators to contemporary haptic platforms, provide repeatable, metric-generating practice (Delorme et al., 2012). Systematic reviews cataloguing this landscape confirm rapid growth: one neurosurgery-focused review and meta-analysis identified 50 distinct simulator types across cranial and spinal applications, encompassing cadaveric, low-fidelity, part-task, and VR systems (Davids et al., 2021), and a companion cranial review reached similar conclusions about breadth (Patel et al., 2020). Focused reviews exist for specific domains, such as the range of training models available for endoscopic transsphenoidal surgery (Santona et al., 2023).

What does the effectiveness evidence show? The foundational demonstration that simulator training can improve subsequent operating-room performance came from a randomized, double-blinded study of VR training in laparoscopic surgery (Seymour et al., 2002), and a systematic review of skills transfer after surgical simulation subsequently concluded that simulation-trained groups generally outperformed conventionally trained comparators on transfer tests (Sturm et al., 2008). A meta-analytic comparative review across health professions found that simulation-based education with deliberate practice was associated with superior outcomes relative to traditional clinical education for a range of skill outcomes (McGaghie et al., 2011). Within neurosurgery, the meta-analysis by Davids and colleagues reported favorable effects of simulation on procedural knowledge and technical skills, while noting the predominance of small studies (Davids et al., 2021). Individual randomized trials add texture: rehearsal on a patient-specific VR platform improved aneurysm clipping performance measures in a prospective randomized study (Chugh et al., 2017), and practice on an augmented-reality haptic simulator improved residents' performance of ventriculostomy (Yudkowsky et al., 2013). At the curricular level, the Society of Neurological Surgeons boot camp courses established a national fundamentals curriculum for entering neurosurgery residents in the United States, normalizing simulation as an induction-level activity (Selden et al., 2012).

Several limitations temper this generally positive picture and are directly relevant to planning in China. First, much of the evidence addresses early skill acquisition in novices and is measured on simulators themselves; evidence of transfer to live neurosurgical performance, and especially to patient outcomes, remains sparse and largely indirect (Davids et al., 2021; Sturm et al., 2008). Second, studies are typically small, single-center, and short in follow-up, and a systematic review of simulation validity evidence found that key sources of validity evidence and methodological

reporting were often incomplete (Cook et al., 2013). Third, model validation is inconsistent: a review evaluating neurosurgical simulation models against face, content, and construct validity criteria found that relatively few models had accumulated comprehensive validity evidence (Chawla et al., 2022). Fourth, higher cost and fidelity do not automatically translate into better learning; the instructional design surrounding the simulator matters more than the device. These caveats argue for a Chinese strategy that adopts simulation confidently for early and mid-stage training while investing in domestic studies of transfer and patient-linked outcomes, rather than assuming benefit.

3.3. Measurement, Feedback, and Learning Design

If simulation supplies the practice environment, assessment supplies its educational discipline. The foundational instrument for structured technical assessment is the Objective Structured Assessment of Technical Skill (OSATS), which combined procedure-specific checklists with global rating scales and demonstrated reliable discrimination of surgical skill (Martin et al., 1997). Neurosurgery has since developed specialty-adapted instruments. The Northwestern Objective Microanastomosis Assessment Tool (NOMAT) showed face and construct validity for microanastomosis performance and was designed for use in modular simulation curricula (Aoun et al., 2015). The Neuro-Endoscopic Ventriculostomy Assessment Tool (NEVAT) accumulated validity evidence for assessing endoscopic third ventriculostomy performance (Breimer et al., 2017). Validity studies of biological models, such as the face, content, and construct validation of a human-placenta aneurysm clipping model, illustrate how the simulation environment itself should be interrogated before being embedded in curricula (Belykh et al., 2017).

Underlying these instruments is contemporary validity theory. Messick's unified framework defines validity as the degree of evidence supporting interpretation and use of assessment scores, encompassing content, response process, internal structure, relations to other variables, and consequences (Messick, 1989). This framework matters practically: a simulator score should not drive entrustment decisions unless evidence supports each of these sources for the intended use. Appraisal tools such as the Medical Education Research Study Quality Instrument (MERSQI) help programs and reviewers judge the methodological quality of educational studies, an increasingly important function as the literature grows (Cook & Reed, 2015). Across the simulation assessment literature, incomplete validity evidence remains the norm rather than the exception (Cook et al., 2013), reinforcing the need for local validation when instruments are adopted in new settings.

Instructional design determines how simulation time is spent. Case-based learning has accumulated systematic-review support in health professions education as a means of linking technical content to clinical reasoning (Thistlethwaite et al., 2012), and comparisons of case-based and problem-based formats suggest differences in faculty workload and learner experience rather than dramatic differences in knowledge outcomes (Srinivasan et al., 2007). Spaced education—delivering content in repeated, time-separated cycles—improved retention of clinical knowledge in randomized trials among medical students and generated transfer of diagnostic skills in surgical contexts (Kerfoot et al., 2007; Kerfoot et al., 2010). Flipped-classroom designs, in which content acquisition moves online and class time is reserved for application, have shown modestly favorable academic outcomes in meta-analyses, though with substantial heterogeneity and study-quality

concerns that warrant conservative interpretation (Chen et al., 2018; Spaic et al., 2025). Feedback is the engine of deliberate practice, and the evidence synthesis on effective feedback in health professions education underscores the importance of specific, credible, and actionable feedback delivered within a supportive relationship (Archer, 2010).

Artificial intelligence (AI) is an emerging adjunct rather than a replacement faculty. In a randomized clinical trial, medical students learning simulated neurosurgical tumor resection with an AI tutoring system achieved performance outcomes comparable in key respects to those receiving remote expert instruction, suggesting that AI-assisted feedback may help scale deliberate practice where expert time is scarce (Fazlollahi et al., 2022). At the same time, recent AMEE guidance on AI in medical education research and on the ethical use of AI in health professions education cautions that such systems should function as assisted feedback with human oversight, transparent limitations, and attention to bias, data governance, and accountability (Masters, 2023; Tolsgaard et al., 2023). For resource-constrained settings, the pragmatic reading is that AI-assisted feedback is a promising way to extend scarce expert capacity, provided that validity evidence is accumulated locally and entrustment decisions remain human responsibilities.

4. From Equipment to System: A Curriculum Implementation Framework

International experience is useful only if it is translated into the governance, resource distribution, and workforce scale of Chinese neurosurgical training. Implementation should therefore proceed as a sequence of system-building tasks rather than as a procurement exercise. During an initial standards-and-pilot stage, specialty educators and training-program leaders could define common procedure-level expectations and test them in a small network of centers that already maintain skills laboratories. A second, distributed-expansion stage would broaden participation by supplying validated low-cost models, preparing regional instructors, and connecting curricula through shared digital resources. A later accountability stage would link simulation benchmarks to workplace entrustment and use multicenter data to refine standards. This sequence makes governance and evidence the drivers of expansion, rather than treating technology deployment as an endpoint.

4.1. Define Procedure-Level Entrustment Expectations

A first task is to define, in Chinese regulatory and educational language, the technical competencies expected at each training stage. International frameworks provide starting points rather than finished products: CBME supplies the outcome-oriented logic (Frank et al., 2010), EPAs supply the mechanism for workplace entrustment (ten Cate, 2005; ten Cate et al., 2015), and the Neurological Surgery Milestones illustrate how developmental trajectories can be structured for a surgical specialty (ACGME & Society of Neurological Surgeons, 2018). Because the milestones reflect United States training structures and are not a certification instrument, direct adoption would be inappropriate; instead, a specialty working group—drawing on professional societies, residency-training representatives, and simulation educators—could define Chinese neurosurgical EPAs for a manageable starter set, such as positioning and draping, ventriculostomy, microanastomosis, basic cranial opening, and endoscopic ventricular procedures. The vascular neurosurgery EPA example

demonstrates that procedure-level entrustment scales are feasible to construct and can be designed explicitly to incorporate simulation-based evidence of readiness (van Lieshout et al., 2023).

Each EPA should specify prerequisite simulation benchmarks: the model on which proficiency is demonstrated, the assessment instrument used, the required performance level, and the number of confirming observations. This is where validated tools enter the curriculum: OSATS-style global rating scales for generic technical skill (Martin et al., 1997), NOMAT for microanastomosis (Aoun et al., 2015), and NEVAT for endoscopic ventriculostomy (Breimer et al., 2017) are candidate instruments whose psychometric properties should be re-examined in Chinese cohorts before high-stakes use, consistent with a Messick-style validity agenda (Messick, 1989). Program accreditation or institutional review could then ask each training center to document the simulation activity, calibrated assessment, and faculty oversight associated with every procedure-level expectation.

4.2. Match Infrastructure Investment to Training Value

Cost is the most frequently cited barrier to simulation adoption, documented in surveys of United States program directors and certain to be amplified in a system training large numbers of residents across heterogeneous institutions (Ganju et al., 2013). A tiered infrastructure model addresses this directly. At the apex, a small number of regional or high-resource simulation hubs—co-located with existing academic skills laboratories—would host high-fidelity resources such as VR platforms with haptics (Delorme et al., 2012), patient-specific 3D-printed rehearsal models (Ryan et al., 2016), and cadaveric facilities, and would serve as centers for faculty development, rater training, and research. At the base, every accredited training base should guarantee access to low-cost, high-yield models: bench microanastomosis stations, placenta-based vascular models that have documented face, content, and construct validity (Belykh et al., 2017), and locally 3D-printed procedural models, a category for which systematic review shows a broad and expanding menu (Santona et al., 2023).

Bootcamp-style concentrated courses offer a second, cost-efficient delivery channel. The Society of Neurological Surgeons boot camp experience shows that a national fundamentals curriculum for entering residents is organizationally feasible (Selden et al., 2012), and a national-level simulation bootcamp developed outside a high-income setting documented both educational delivery and cost structure in detail, providing a realistic planning reference for Chinese organizers (Bakhshi et al., 2022). A sensible Chinese sequence is to introduce a standardized induction bootcamp for new neurosurgery residents covering microsurgical fundamentals, followed by procedure-specific modules at defined training stages. Procurement should follow validity evidence rather than fidelity for its own sake: the review by Chawla and colleagues makes clear that relatively few models have accumulated comprehensive validity evidence, so purchasing decisions should privilege instruments and models with published validation and should budget for consumables and maintenance, not only capital equipment (Chawla et al., 2022).

4.3. Build a Distributed Instructor and Rater Network

Simulation programs fail more often from faculty scarcity than from equipment scarcity. Reliable assessment requires calibrated raters; the validation pathways of NOMAT and NEVAT both depended on deliberate rater training and reliability testing (Aoun et al., 2015; Breimer et al., 2017). A structured train-the-trainer program should therefore accompany infrastructure rollout, certifying

simulation instructors in three domains: curriculum delivery (briefing, deliberate practice management, debriefing), assessment (use of rating instruments, calibration exercises, documentation of entrustment recommendations), and feedback skills. The evidence synthesis on feedback in health professions education emphasizes that effective feedback is specific, credible, and delivered within an educational alliance—skills that are teachable and assessable (Archer, 2010).

Train-the-trainer structures also address regional disparity. Hub centers can prepare regional faculty who then deliver standardized curricula locally, with periodic recalibration. AI-assisted feedback may extend this model where expert time is scarcest: the randomized trial of an AI tutor for simulated neurosurgical resection suggests that automated performance feedback can approximate expert remote instruction for defined tasks (Fazlollahi et al., 2022). Adoption should nonetheless follow the boundaries articulated in recent AMEE guidance—human oversight of consequential decisions, transparency about system limitations, attention to bias and data protection, and local validation before reliance (Masters, 2023; Tolsgaard et al., 2023). In practice, this means positioning AI as a first-pass feedback assistant for routine simulator metrics, with faculty retaining responsibility for entrustment decisions and for coaching judgment, situational awareness, and non-technical performance.

4.4. Connect Digital Learning to Shared Evaluation

Digital platforms can distribute both content and standards across China's geography. Flipped-classroom designs—online acquisition of procedural knowledge followed by laboratory time devoted to guided practice—fit simulation curricula naturally and have meta-analytic support, albeit with heterogeneity that counsels conservative expectations (Chen et al., 2018; Spaic et al., 2025). Spaced education, delivered by email or learning platforms in repeated cycles, improved knowledge retention in randomized trials and is inexpensive to scale across institutions (Kerfoot et al., 2007; Kerfoot et al., 2010). Case-based learning can connect technical modules to clinical decision-making and has systematic-review support in health professions education (Thistlethwaite et al., 2012); a single-institution Chinese study has reported teaching reform using case-based and problem-based methods in neurosurgery (Shen et al., 2022), although one such report cannot establish the wider receptivity of programs to these formats. Two cautions apply. First, technology is a delivery channel, not an educational effect: systematic reviews of technology-enhanced simulation emphasize that outcomes depend on instructional design and that reporting quality in this literature is uneven (Cook et al., 2013). Second, blended designs should be evaluated against the same standards as face-to-face teaching rather than assumed superior because they are digital.

Evaluation should be planned before rollout. Kirkpatrick's four-level model—reaction, learning, behavior, and results—offers a familiar evaluation scaffold, but educational research frequently misapplies it by treating satisfaction as evidence of effectiveness; pairing the model with the critical interpretation offered by Yardley and Dornan helps programs set realistic expectations about what each level can and cannot show (Kirkpatrick, 1994; Yardley & Dornan, 2012). Study quality should be appraised with instruments such as MERSQI (Cook & Reed, 2015). Reporting should follow established guidelines so that Chinese studies are interpretable and cumulative: TIDieR for describing interventions (Hoffmann et al., 2014), GREET for educational interventions (Phillips et

al., 2014; Phillips et al., 2016), DoCTRINE for reporting innovations in education (Blanco et al., 2022), STROBE for observational educational studies (von Elm et al., 2007), and the CONSORT statement together with its non-pharmacological extension for the randomized trials of simulation curricula that the field needs (Boutron et al., 2008; Schulz et al., 2010). A shared simulation registry—recording curricula, instruments, benchmarks, and outcomes across pilot centers—would accelerate both quality improvement and research, and would allow China to test the central unanswered question in this literature: whether simulation-based training changes operative performance and patient outcomes in its own training system, rather than relying on indirect evidence from other countries and other specialties (Davids et al., 2021; McGaghie et al., 2011; Sturm et al., 2008).

5. Discussion

Several constraints should qualify these recommendations. First, this article is a narrative synthesis rather than a national audit. We identified no comprehensive province-level data describing the availability, use, staffing, or educational outcomes of neurosurgical simulation in China, so the present account may miss important local innovation as well as local weakness. Second, a substantial part of the international effectiveness literature comes from other surgical specialties, novice participants, or endpoints measured on simulators; extrapolation to Chinese neurosurgical residents and to clinical performance should remain cautious. Third, many candidate assessment tools have been studied in small cohorts and in single national contexts, and their measurement properties cannot simply be assumed to transfer across languages, training cultures, and case mixes. The next research agenda is therefore practical as well as scholarly: map existing resources and practices through a national survey or registry; validate candidate rating instruments in Chinese programs; compare implementation strategies in multicenter educational trials; examine transfer to supervised operative performance; and estimate the costs and consequences of different infrastructure models. These studies would allow Chinese educators to refine the proposed curriculum with domestic evidence rather than relying indefinitely on imported assumptions.

6. Conclusions

The strongest case for simulation in neurosurgery is not that a device can imitate an operation. It is that a well-governed simulation curriculum can make technical development visible before patients are exposed to the steepest part of the learning curve. International studies support the use of structured practice for early skill acquisition, but they also show why equipment alone is insufficient: learning gains depend on explicit competency targets, appropriate model selection, trained instructors, credible feedback, and assessment scores backed by validity evidence. China already possesses an important starting point in a national training pathway that includes a dedicated microsurgical-skills rotation (Xu et al., 2016). The next step is to connect that foundation to procedure-level entrustment expectations, infrastructure chosen for educational yield, a distributed instructor network, and digital systems that support evaluation rather than merely content delivery. Such a program would convert simulation from a local training accessory into a systematic

mechanism for safer and more consistent neurosurgical education. Its success should ultimately be judged not by the number of simulators purchased, but by whether trainees progress more reliably, faculty make better-documented entrustment decisions, and multicenter evidence demonstrates transfer into supervised operative performance.

Author Contributions:

Conceptualization, H.X. and Z.Z.; methodology, C.M. and C.L.; writing ─ original draft preparation, H.X.; writing─review and editing, X.H. and D.Z. All authors have read and agreed to the published version of the manuscript.

Funding:

This research was funded by the National Natural Science Foundation of China, grant number 82271396.

Institutional Review Board Statement:

Not applicable.

Informed Consent Statement:

Not applicable.

Data Availability Statement:

No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflict of Interest:

The authors declare no conflict of interest.

Reference

Accreditation Council for Graduate Medical Education, & Society of Neurological Surgeons. (2018). The Neurological Surgery Milestones. https://www.acgme.org/globalassets/pdfs/milestones/neurologicalsurgerymilestones.pdf

Aoun, S. G., El Ahmadieh, T. Y., El Tecle, N. E., et al. (2015). A pilot study to assess the construct and face validity of the Northwestern Objective Microanastomosis Assessment Tool. Journal of Neurosurgery, 123(1), 103–109. https://doi.org/10.3171/2014.12.JNS131814

Archer, J. C. (2010). State of the science in health professional education: effective feedback. Medical Education, 44(1), 101–108. https://doi.org/10.1111/j.1365-2923.2009.03546.x

Bakhshi, S. K., Ahmad, R., Merchant, A. A. H., et al. (2022). Development, outcome and costs of a simulation-based neurosurgery bootcamp at the national level. BMC Medical Education, 22, 896. https://doi.org/10.1186/s12909-022-03965-9

Belykh, E., Miller, E. J., Lei, T., Chapple, K., Byvaltsev, V. A., Spetzler, R. F., Nakaji, P., & Preul, M. C. (2017). Face, content, and construct validity of an aneurysm clipping model using

human placenta. World Neurosurgery, 105, 952–960.e2. https://doi.org/10.1016/j.wneu.2017.06.045

Blanco, M., Prunuske, J., DiCorcia, M., Learman, L. A., Mutcheson, B., & Huang, G. C. (2022). The DoCTRINE guidelines: Defined Criteria To Report INnovations in Education. Academic Medicine, 97(5), 689–695.

Boutron, I., Moher, D., Altman, D. G., Schulz, K. F., Ravaud, P., & CONSORT Group. (2008). Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Annals of Internal Medicine, 148(4), 295–309.

Breimer, G. E., Haji, F. A., Cinalli, G., Hoving, E. W., & Drake, J. M. (2017). Validity evidence for the neuro-endoscopic ventriculostomy assessment tool (NEVAT). Operative Neurosurgery, 13(1), 60–68. https://doi.org/10.1227/NEU.0000000000001158

Chawla, S., Devi, S., Calvachi, P., Gormley, W. B., & Rueda-Esteban, R. (2022). Evaluation of simulation models in neurosurgical training according to face, content, and construct validity: a systematic review. Acta Neurochirurgica, 164(4), 947–966. https://doi.org/10.1007/s00701-021-05003-x

Chen, K. S., Monrouxe, L., Lu, Y. H., Jenq, C. C., Chang, Y. J., Chang, Y. C., & Chai, P. Y. C. (2018). Academic outcomes of flipped classroom learning: a meta-analysis. Medical Education, 52(9), 910–924. https://doi.org/10.1111/medu.13616

Chugh, A. J., Pace, J. R., Singer, J., Tatsuoka, C., Hoffer, A., Selman, W. R., & Bambakidis, N. C. (2017). Use of a surgical rehearsal platform and improvement in aneurysm clipping measures: results of a prospective, randomized trial. Journal of Neurosurgery, 126(3), 838–844. https://doi.org/10.3171/2016.1.JNS152576

Cook, D. A., Brydges, R., Zendejas, B., Hamstra, S. J., & Hatala, R. (2013). Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. Academic Medicine, 88(6), 872–883. https://doi.org/10.1097/ACM.0b013e31828ffdcf

Cook, D. A., & Reed, D. A. (2015). Appraising the quality of medical education research methods: the Medical Education Research Study Quality Instrument and the Newcastle-Ottawa Scale-Education. Academic Medicine, 90(8), 1067–1076. https://doi.org/10.1097/ACM.0000000000000786

Davids, J., Manivannan, S., Darzi, A., Giannarou, S., Ashrafian, H., & Marcus, H. J. (2021). Simulation for skills training in neurosurgery: a systematic review, meta-analysis, and analysis of progressive scholarly acceptance. Neurosurgical Review, 44(4), 1853–1867. https://doi.org/10.1007/s10143-020-01378-0

Delorme, S., Laroche, D., DiRaddo, R., & Del Maestro, R. F. (2012). NeuroTouch: a physics-based virtual simulator for cranial microneurosurgery training. Neurosurgery, 71(1 Suppl Oper), 32–42.

Fazlollahi, A. M., Bakhaidar, M., Alsayegh, A., et al. (2022). Effect of artificial intelligence tutoring vs expert instruction on learning simulated surgical skills among medical students: a randomized clinical trial. JAMA Network Open, 5(2), e2149008. https://doi.org/10.1001/jamanetworkopen.2021.49008

Frank, J. R., Snell, L. S., Ten Cate, O., Holmboe, E. S., Carraccio, C., Swing, S. R., et al. (2010). Competency-based medical education: theory to practice. Medical Teacher, 32(8), 638–645.

Ganju, A., Aoun, S. G., Daou, M. R., et al. (2013). The role of simulation in neurosurgical education: a survey of 99 United States neurosurgery program directors. World Neurosurgery, 80(5), e1–e8. https://doi.org/10.1016/j.wneu.2012.11.066

Hoffmann, T. C., Glasziou, P. P., Boutron, I., et al. (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ, 348, g1687. https://doi.org/10.1136/bmj.g1687

Kerfoot, B. P., DeWolf, W. C., Masser, B. A., Church, P. A., & Federman, D. D. (2007). Spaced education improves the retention of clinical knowledge by medical students: a randomised controlled trial. Medical Education, 41(1), 23–31. https://doi.org/10.1111/j.1365-2929.2006.02644.x

Kerfoot, B. P., Fu, Y., Baker, H., Connelly, D., Ritchey, M. L., & Genega, E. M. (2010). Online spaced education generates transfer and improves long-term retention of diagnostic skills: a randomized controlled trial. Journal of the American College of Surgeons, 211(3), 331–337.e1. https://doi.org/10.1016/j.jamcollsurg.2010.04.023

Kirkpatrick, D. L. (1994). Evaluating training programs: the four levels. Berrett-Koehler.

Martin, J. A., Regehr, G., Reznick, R., MacRae, H., Murnaghan, J., Hutchison, C., & Brown, M. (1997). Objective structured assessment of technical skill (OSATS) for surgical residents. British Journal of Surgery, 84(2), 273–278. https://doi.org/10.1046/j.1365-2168.1997.02502.x

Masters, K. (2023). Ethical use of artificial intelligence in health professions education: AMEE Guide No. 158. Medical Teacher, 45(6), 574–584. https://doi.org/10.1080/0142159X.2023.2186203

McGaghie, W. C., Issenberg, S. B., Cohen, E. R., Barsuk, J. H., & Wayne, D. B. (2011). Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. Academic Medicine, 86(6), 706–711. https://doi.org/10.1097/ACM.0b013e318217e119

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13–103). American Council on Education and Macmillan.

Patel, E. A., Aydin, A., Cearns, M., Dasgupta, P., & Ahmed, K. (2020). A systematic review of simulation-based training in neurosurgery, part 1: cranial neurosurgery. World Neurosurgery, 133, e850–e873. https://doi.org/10.1016/j.wneu.2019.08.262

Phillips, A. C., Lewis, L. K., McEvoy, M. P., et al. (2014). A systematic review of how studies describe educational interventions for evidence-based practice: stage 1 of the development of a reporting guideline. BMC Medical Education, 14, 152. https://doi.org/10.1186/1472-6920-14-152

Phillips, A. C., Lewis, L. K., McEvoy, M. P., et al. (2016). Development and validation of the guideline for reporting evidence-based practice educational interventions and teaching (GREET). BMC Medical Education, 16, 237. https://doi.org/10.1186/s12909-016-0759-1

Ryan, J. R., Almefty, K. K., Nakaji, P., & Frakes, D. H. (2016). Cerebral aneurysm clipping surgery simulation using patient-specific 3D printing and silicone casting. World Neurosurgery, 88, 175–181. https://doi.org/10.1016/j.wneu.2015.12.102

Santona, G., Madoglio, A., Mattavelli, D., Rigante, M., Ferrari, M., Lauretti, L., et al. (2023). Training models and simulators for endoscopic transsphenoidal surgery: a systematic review. Neurosurgical Review, 46(1), 248. https://doi.org/10.1007/s10143-023-02149-3

Schulz, K. F., Altman, D. G., Moher, D., & CONSORT Group. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ, 340, c332.

Selden, N. R., Origitano, T. C., Burchiel, K. J., Getch, C. C., Anderson, V. C., et al. (2012). A national fundamentals curriculum for neurosurgery PGY1 residents: the 2010 Society of Neurological Surgeons boot camp courses. Neurosurgery, 70(4), 971–981. https://doi.org/10.1227/NEU.0b013e31823d7a45

Seymour, N. E., Gallagher, A. G., Roman, S. A., O'Brien, M. K., Bansal, V. K., Andersen, D. K., & Satava, R. M. (2002). Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Annals of Surgery, 236(4), 458–464. https://doi.org/10.1097/00000658-200210000-00008

Shen, J., Yuan, L., Ge, R., Shao, X., & Jiang, X. (2022). Improving medical student recruitment into neurosurgery through teaching reform. BMC Medical Education, 22, 656. https://doi.org/10.1186/s12909-022-03722-y

Spaic, D., Bukumiric, Z., Rajovic, N., Markovic, K., Savic, M., Milin-Lazovic, J., et al. (2025). The flipped classroom in medical education: systematic review and meta-analysis. Journal of Medical Internet Research, 27, e60757. https://doi.org/10.2196/60757

Srinivasan, M., Wilkes, M., Stevenson, F., Nguyen, T., & Slavin, S. (2007). Comparing problem-based learning with case-based learning: effects of a major curricular shift at two institutions. Academic Medicine, 82(1), 74–82. https://doi.org/10.1097/01.ACM.0000249963.93776.aa

Sturm, L. P., Windsor, J. A., Cosman, P. H., Cregan, P., Hewett, P. J., & Maddern, G. J. (2008). A systematic review of skills transfer after surgical simulation training. Annals of Surgery, 248(2), 166–179. https://doi.org/10.1097/SLA.0b013e318176bf24

ten Cate, O. (2005). Entrustability of professional activities and competency-based training. Medical Education, 39(12), 1176–1177. https://doi.org/10.1111/j.1365-2929.2005.02341.x

ten Cate, O., Chen, H. C., Hoff, R. G., Peters, H., Bok, H., & van der Schaaf, M. (2015). Curriculum development for the workplace using entrustable professional activities (EPAs): AMEE Guide No. 99. Medical Teacher, 37(11), 983–1002. https://doi.org/10.3109/0142159X.2015.1060308

Thistlethwaite, J. E., Davies, D., Ekeocha, S., Kidd, J. M., MacDougall, C., Matthews, P., et al. (2012). The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. Medical Teacher, 34(6), e421–e444. https://doi.org/10.3109/0142159X.2012.680939

Tolsgaard, M. G., Pusic, M. V., Sebok-Syer, S. S., Gin, B., Svendsen, M. B., Syer, M. D., et al. (2023). The fundamentals of artificial intelligence in medical education research: AMEE Guide No. 156. Medical Teacher, 45(6), 565–573. https://doi.org/10.1080/0142159X.2023.2180340

van Lieshout, J. H., Malzkorn, B., Steiger, H.-J., Karadag, C., Kamp, M. A., Vajkoczy, P., et al. (2023). Defining activities in neurovascular microsurgery training: entrustable professional

activities for vascular neurosurgery. Acta Neurochirurgica, 165(1), 27–37. https://doi.org/10.1007/s00701-022-05372-x

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., Vandenbroucke, J. P., & STROBE Initiative. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. The Lancet, 370(9596), 1453–1457. https://doi.org/10.1016/S0140-6736(07)61602-X

Xu, T., Evins, A. I., Lin, N., Chang, J., Hu, G., Hou, L., et al. (2016). Neurosurgical postgraduate training in China: moving toward a national training standard. World Neurosurgery, 96, 410–416. https://doi.org/10.1016/j.wneu.2016.09.034

Yardley, S., & Dornan, T. (2012). Kirkpatrick's levels and education 'evidence'. Medical Education, 46(1), 97–106. https://doi.org/10.1111/j.1365-2923.2011.04076.x

Yudkowsky, R., Luciano, C., Banerjee, P., Schwartz, A., Alaraj, A., Lemole, G. M., et al. (2013). Practice on an augmented reality/haptic simulator and library of virtual brains improves residents' ability to perform a ventriculostomy. Simulation in Healthcare, 8(1), 25–31. https://doi.org/10.1097/SIH.0b013e3182662c69

License: Copyright (c) 2026 Author.
All articles published in this journal are licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are properly credited. Authors retain copyright of their work, and readers are free to copy, share, adapt, and build upon the material for any purpose, including commercial use, as long as appropriate attribution is given.