Balancing Scaffolding and Autonomy: A Case Study in Designing a Scalable Undergraduate Machine Learning Research Course

Xinyue Chen, Sharon Jessica, Xu Wang, Sindhu Kutty
University of Michigan, Ann Arbor, MI, 48109, USA
xinyuech@umich.edu, sharjess@umich.edu, xwanghci@umich.edu, skutty@umich.edu

Abstract

Undergraduate research experiences are often limited to small-scale apprenticeship models, leaving many students without accessible entry points into research practice. This paper presents the design and evaluation of a semester-long course for undergraduates to gain research experience in Machine Learning. The course, led by one faculty instructor, enables nearly a hundred students to engage in structured research through a scaffolded replication-and-extension project, where students first replicate a published research project and then implement novel additions. The course integrates instructional modules (e.g., guided paper reading, proposal writing, public presentation) with project milestones (e.g., replication, extension, poster) to support research learning for students with diverse backgrounds. Every component of research is visited several times, with each iteration having progressively increased autonomy coupled with simultaneously decreased scaffolding. We find that the scaffolding modules help students develop foundational conceptual and procedural understanding of doing research, and the project milestones on replication and extension help them gain execution skills gradually. Students also report developing a researcher mindset and feeling like they understand the research process better. We discuss the principles used to design a scalable research-based course: balancing *scaffolding* to provide foundational understanding with *autonomy* for students to "feel like real researchers".

Introduction

Undergraduate research experiences (UREs) foster critical thinking, technical skills, and the development of a scientific identity (Linn et al. 2015; Stephens, Brenner, and Gentile 2017; Bangera and Brownell 2014). However, UREs are traditionally structured around one-on-one faculty mentorship, making them difficult to scale within large computer science departments. Course-Based Undergraduate Research Experiences (CUREs) emerged as a scalable alternative, offering inclusive and structured pathways into research through the classroom (Linn et al. 2015). While CUREs have gained traction in disciplines like biology and chemistry, their uptake in Computer Science (CS) remains limited (Fernandez 2025; Alvarado, Villazon, and Tamer 2019). This is especially concerning given the rapid growth of CS programs— our own department enrolls over 3,000 undergraduates — and the corresponding strain on instructional resources. Moreover, CS is a field well-suited to research-based learning, where hands-on experimentation and iterative design mirror authentic practice.

To address this gap, we have designed Machine Learning Research Experience (MLRE), a semester-long course at a large public R1 university that guides students through the full research process—from literature review and idea inception to implementation and evaluation. The course enrolls over 90 undergraduates across four CS-related programs, all supported by a single faculty instructor. MLRE uses a scaffolded replication-and-extension model: students begin by reproducing results from a published ML paper, then propose and implement novel extensions grounded in literature analysis. The course structure is modular and iterative, with students engaging in cycles of paper reading, technical writing, replication, and presentation, each revisited with decreasing scaffolding and increasing autonomy. Students work in teams and move through a coordinated set of instructional modules: guided paper reading, structured proposal writing, project replication and extension, peer-reviewed presentation rounds, and a final mock conference including paper submission, review, rebuttal, and poster presentation. Scalability is achieved through built-in support rather than additional mentors. These support structures include assignment templates and structured peer feedback. Check-in points with the instructional faculty member are strategically aligned with key milestones.

In order to understand how effectively the course supported research learning, we conducted post-course focus groups with 26 students who opted into attending the focus groups. Students described how lectures and research tasks, structured paper reading, and summaries built their foundational understanding of research framing. The replication-extension model provided a gentler entry, helping students to transition from understanding prior work to contributing novel ideas. Students also identified areas where scaffolds were misaligned or insufficient. Some students noted their discomfort during the less scaffolded parts of the course, for instance, indicating a preference for a stronger connection between early readings and their ultimate choice of project, and struggling with the open-ended nature of the extension phase. These reflections suggest opportunities for future refinement, such as supplementary targeted reading and more check-ins during the extension phase. Despite these tensions, students widely valued the autonomy-with-structure model. Even with limited instructional staff, the layered scaffolding enabled students to engage meaningfully and develop confidence as novice researchers. These findings affirm the MLRE course design as a scalable approach to equitable undergraduate research in CS and offer design implications for future CURE models that better align structure, autonomy, and support for diverse learners.

Related Work and Context

Course-Based Undergraduate Research Experiences (CUREs) and Their Application in Computer Science

Course-Based Undergraduate Research Experiences (CUREs) offer a scalable and inclusive alternative to traditional mentored research opportunities (UREs), which — while impactful for student learning and identity development— remain resource-intensive and limited in availability (Bangera and Brownell 2014; Granger et al. 2006). By embedding inquiry-driven, collaborative research into the curriculum, CUREs seek to democratize access and broaden participation in undergraduate research (Bekkering 2025). In disciplines such as biology and chemistry, CUREs have become well-established and supported by infrastructure like CUREnet (Connors et al. 2021). However, in computer science (CS), CUREs remain rare, often limited to small-scale projects within existing courses (Bekkering 2025). While some CS pedagogy includes open-ended components like design projects or capstones, they often lack the ambiguity and iteration of authentic research. Recently, a few courses offer sustained training in core research practices (Fernandez 2025; Dougherty 2024). Those explorations found that without sufficient scaffolding, students often struggle with scoping and navigating the uncertainty of inquiry (Bekkering 2025), exposing a deeper epistemic gap between traditional CS coursework and real research (Bekkering 2025). Meanwhile, sub-fields like machine learning (ML) offer unique opportunities for integrating research into undergraduate education: open-source code, public datasets, and standardized benchmarks make replication and extension feasible entry points (Werner et al. 2024). Embedding these practices into the curriculum can provide a scalable and timely response to growing student interest in AI and research more broadly.

Scaffolding as a Pedagogical Bridge from Instruction to Autonomy

Scaffolding provides structured support early in complex tasks, using tools like templates and checklists to help students navigate unfamiliar challenges (Belland and Belland 2017). In CS research-oriented courses, scaffolding often takes the form of staged project components with timelines, rubrics, and feedback. For example, one CS research course initially saw low student progress due to minimal structure, but later improved outcomes by introducing milestone deadlines and planning templates (Dougherty 2024). As students gain competence, fading scaffolds gradually reduce support, intentionally shifting cognitive responsibility to the learner (McNeill et al. 2006). When teaching research skills, carefully sequenced and faded scaffolds can help ensure that students acquire key research skills in a structured way, while still fostering the autonomy needed for authentic, open-ended research (Vygotsky and Cole 1978; Russell, Hancock, and McCullough 2007). We deliberately applied this strategy in our course to balance structured support with increasing student ownership.

Institutional Context and Motivation

Our course was implemented in a large public R1 university, where the Division of Computer Science enrolls over 3,000 undergraduate students across multiple degree tracks in both the Engineering and Literature, Science and Arts colleges. With the low faculty-to-student ratio, traditional pathways, like faculty-mentored projects, or university-wide programs UROP, are highly competitive and limited in capacity (Goodlad 1998).

To address this gap, we first piloted a small summer research program where a single faculty member mentored 20 undergraduates on machine learning research projects. Despite limited instructional resources, the pilot showed that with appropriate structure, undergraduates could engage in research at scale (Kutty and Guzdial 2020). Building on this model, we created a semester-long course — Machine Learning Research Experience (MLRE)—taught by a single faculty instructor and enrolling more than 90 students. The course integrates instructional modules with project milestones. The design and implementation of MLRE demonstrates that structured, authentic research experiences can be equitably and scalably delivered to undergraduates, even in resource-constrained settings.

Course Structure and Implementation

To support students through the entire life-cycle of machine learning research projects at scale, the course introduces and integrates two parallel components: 1) Carefully designed general-purpose modules on foundational research skills, e.g., guided paper reading, technical writing, and public presentation. Each of these modules is visited several times throughout the semester. 2) A scaffolded replication-and-extension project where students gain hands-on experience replicating a published ML paper of their choice, and then propose and evaluate an extension to the original work. Students are expected to gain authentic research experience in the sub-area of their choosing. The open-ended projects and the foundational modules are tightly coupled in pacing and focus. A detailed course module timeline and assignment samples can be shared as needed.

Course Structure and Module Overview

Foundational Modules on General Research Skills (Lectures and Guided Activities) The lecture sequence is designed to equip students with the necessary research mindset and skills in sync with project milestones. Early lectures

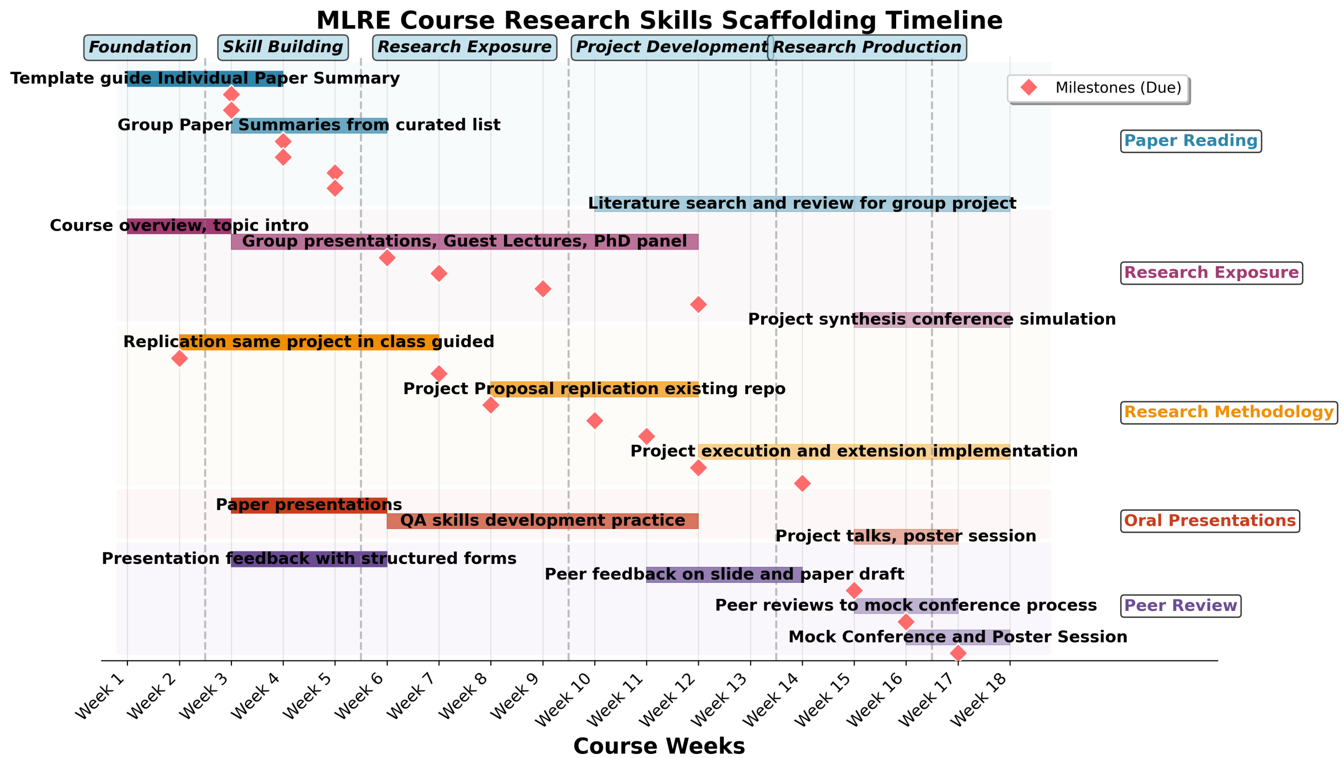

Figure 1: **Layered Scaffolds for Different Research Skills.** Each row represents a major research skill (e.g., paper reading, research methodology, oral presentation), and the colored bars reflect the extent and duration of instructor-provided scaffolds for that skill. **Darker shades** on each bar indicate intensive scaffolding with structured guidance (e.g., templates, instructor modeling), while **lighter shades** represent faded support as student autonomy increases. Student deliverables (marked in red diamonds) align with these shifts, requiring the timely application of evolving competencies. The course thus supports a gradual transition from supported practice to independent execution.

focus on research framing—how to identify research questions, reason about contributions, understand methodological choices, and evaluate results. This is achieved through a broad literature survey that includes a mix of classical and recent papers. The paper reading component is scaffolded via a paper reading template (provided as supplementary material) that students fill in – first individually and then as a group. Mid-semester sessions shift to modeling proposal writing, reading papers within the area of students' choosing, and surfacing design tradeoffs. In this phase, students are required to meet in groups and record their meeting minutes in a template provided (see supplemental material). Later sessions address how to interpret replication results, propose extensions, and communicate work effectively. Throughout, students are exposed to diverse ML domains through curated papers, group presentations, and guest talks. Rather than being standalone lectures, these sessions combine live demos, guided analysis, and in-class synthesis.

A Semester-Long Replication-and-Extension Project Students work in groups to complete a semester-long project that mirrors a typical ML research pipeline. This includes:

- A **proposal** stage, where students scope a feasible project anchored in a prior paper of their choice.
- A **replication** phase, supported by an initial warm-up task and biweekly instructor check-ins.
- An **extension** phase, where students design and test novel variations or additions to the replicated work.
- A **communication** phase, involving a research-style paper (submitted on the Easy Chair conference management system), a lightning talk, and a public poster presentation (assignments provided as supplementary materials).

Each phase is scaffolded by lecture content, assignment templates, and feedback from instructors and peers.

Exposure and Reflection Opportunities To broaden students' understanding of research practice and pathways, the course includes multiple guest lectures from ML researchers and a PhD student panel. These sessions give students insight into real-world challenges, workflows, and career considerations, reinforcing the authenticity of their own work. Students are required to reflect on each of these sessions and submit these reflections as an assignment.

Manifested Instructional Design Principles and Learning Sciences Theories in MLRE

These three strands—foundational skills instruction, scaffolded project execution, and authentic exposure—are not delivered in isolation. Instead, they are interleaved across the semester to provide timely support and reinforce core competencies through repeated, varied practice.

Their educational value comes from a coherent scaffolding design that integrates the strands: each module is underpinned by instructor-led, material-based, and peer-based scaffolds that align with course milestones and fade from hands-on to hands-off as students gain competence. Figure 1 depicts the course timeline and the planned fading of support, while Table 1 details how scaffolds are organized—categorizing them (instructor/material/peer), specifying their targeted functions and moments of use, and summarizing evidence of effect as well as student-identified frictions and adjustments.

We explained some instructional choices that draw from theory-driven principles below.

Blending Direct Instruction with Project-Based Learning. Prior research has shown that direct instruction helps students acquire knowledge and skills on defined objectives(Klahr and Nigam 2004), whereas project-based learning helps increase student motivation, engagement, and develop identity (Carrabba and Farmer 2018). The course design aims to amplify the benefits of both worlds.

Deliberate and Repeated Practice. Moreover, in the foundational research skill modules, we also introduce repeated practice opportunities(Anders Ericsson 2008). Core research skills are revisited across multiple formats and contexts throughout the course. For example, writing and communication are reinforced through early summaries, peer presentations, and a final full-length paper. These repeated, feedback-rich opportunities embody the principles of deliberate practice (Anders Ericsson 2008)—offering structured repetition with increasing complexity, which can build knowledge fluency and transfer.

Fading Scaffolds to Support Progressive Autonomy A central instructional principle in MLRE is the strategic fading of scaffolds to support students' transition from guided practice to independent research (Martin et al. 2019; Ambrose et al. 2010; Vygotsky and Cole 1978). The course is designed so that students initially receive structured support (e.g., templates, and prompts), which is gradually removed as they gain competence (McNeill et al. 2006).

For example, there are three rounds of guided paper reading in the course, with faded support. Students begin with a shared paper selected by the instructor, accompanied by a structured template that guides them through identifying research motivation, methodology, and key contributions. The template facilitates active learning rather than passive reading (Chi and Wylie 2014; Lu et al. 2023), which can be applied to paper reading beyond this course. In the second round, students select papers from a curated list, write individual summaries, and collaborate on a group presentation—scaffolding their ability to synthesize and communicate literature. In the final round, scaffolding is further reduced: students choose relevant papers for their project area and are expected to identify critical points of discussion independently. A similar progression occurs in project execution. Students first complete a scaffolded replication of a shared ML paper to learn the basic replication skills, then design and implement a replication-extension project among the group with less scaffolds. For a detailed view of how scaffolds evolve across different research skills, see in supplementary material.

Reproducibility as an Epistemic Entry Point in ML The course structure is also informed by disciplinary norms in machine learning, where reproducibility is central to credible research (Pineau 2020; Hutson 2018). Recent work in CS and ML education has similarly adopted reproducibility as a structured learning method—using replication-and-extension assignments to help students practice rigorous, transparent inquiry and understand research as a collective process (Fund 2023; Lucic et al. 2022; Yildiz et al. 2021). Building on this pedagogical framing, our course adopts reproducibility as the entry point for research learning. Students begin the project sequence by replicating a published ML paper using publicly available code and data. This design lowers the barrier to entry and immerses students in authentic artifacts. Replication also serves as a cognitive bridge to indipendent work on extension: by grounding their understanding in existing work, students are better equipped to scope feasible, well-justified extensions.

Design for Scale

The course design also emphasizes scalable instructional strategies that enable 90+ students to engage in research without relying on intensive one-on-one mentorship.

- **Embedded scaffolds reduce one-on-one load:** Much of the support is built into the course structure itself—through templates, rubrics, peer review, and staged deliverables, minimizing the need for continuous individual attention.
- **Strategic use of feedback cycles:** Instructor check-ins are limited but high-leverage, occurring at key decision points (proposal, replication, extension). Peer feedback supplements instructor time, especially during the reading and presentation phases.
- **Clear role division across staff:** The faculty instructor focuses on concept framing and milestone pacing. The faculty instructor also meets with students in small groups at the start of the semester, before each group presentation and provides intermittent check-ins during the proposal and replication phase of the project. The TA manages grading and replication support. The IA handles logistics, communication, and student coordination.
- **Low-cost structures for monitoring progress:** Required group notes and milestone submissions make progress visible without requiring intensive oversight. This allows timely intervention only when necessary.
- **Flexibility for heterogeneous teams:** The project structure allows different groups to specialize based on

strengths - some push on experimental novelty, others on replication depth or communication clarity. The grading rubric is designed to reward thoughtful execution across these dimensions.

Study on Student Experience in MLRE

The Machine Learning Research Experience (MLRE) course ran in Fall 2024 with over 90 students, primarily senior Computer Science majors, alongside Data Science majors and CS minors from other departments.

To understand students' experiences, we conducted a post-course qualitative study during Winter 2025 (after the conclusion of the course), holding 6 IRB-approved focus groups with 26 students from the course, each group with 4-6 students. Using a shared discussion protocol, we investigated students' perceptions of research learning, the scaffolding of research, their autonomy, and the course's career influence. Thematic analysis of transcripts revealed how course components supported research skill development, surfaced key challenges, enabled scalable mentorship, and shaped students' research identities and future goals.

Findings

How Does the Design of Course Modules Scaffold Students' Learning of the Research Process?

We found that the course design effectively scaffolded students' understanding of core research practices through structured modules and iterative support.

Building Conceptual and Procedural Foundations of Research through Lectures and Guided Reading In the early weeks of the course, students often found themselves grappling with a basic but fundamental question: "What is machine learning research?" Through lectures, guided paper readings, and structured summaries, they began to build a shared conceptual map of the field. *"Early on… a good part of the class is getting like a breadth… formalizing like a map of the domains."* (P06)

Students began by reading and summarizing research papers using a structured format, which helped them understand research organization and key focus areas. *" It's pretty easy to read a paper I haven't seen before… and also determine if it's like a good or bad paper."* (P03)

Students contrasted MLRE with prior implementation-focused courses, noting a shift toward understanding the intent behind experimental design, not just outputs: *"I started to think about why someone would design an experiment this way, or whether it actually proves the point."* (P10)

By the proposal stage, students had built confidence and were able to leverage paper structures and contribution types taught in the course to frame their ideas. *"We had to figure out what counts as a good idea… it made us commit to something concrete."* (P01)

Supporting Research Execution through Project Milestones, Team Collaboration, and Feedback Cycles Students often began with vague or overly ambitious ideas. The structured proposal milestone prompted them to define objectives, methods, and roles, pushing for clarity: *"It really made us get like a concrete idea of how the project was going to go, what everyone's part in that was."* (P06) Weekly check-ins and feedback cycles helped refine feasibility: *"She gave really detailed feedback… helped us identify areas we had to flesh out or reel back."* (P09)

Students consistently described the **replication phase** as a critical turning point in understanding how research works. *"Having the replication part is like the groundwork for understanding what's actually going on after reading the papers."* (P12) Replication offered a structured entry point as students were given a concrete system to unpack, understand, and rebuild. *"If we had to just come up with a completely new idea, it's like out of the air,. But since we already had a paper and a setup, we know what to do"* (P5)

The **extension phase** emerged naturally from this process, with students proposing local changes and reasoning about their impact: *"We were already deep in the code, so thinking of what to change wasn't like coming up with an idea from scratch—it was more like seeing something and thinking, could we try this differently?"* (P13) The progressive autonomy allowed students to navigate real research tradeoffs. *"We just changed one part of the evaluation, but we had to defend why."* (P01)

Students were also developing collaborative coordination skills critical to executing shared work. They emphasized the importance of staying connected with the team to the full research logic, *"You have to work on different portions. You can't just be assigned one part."* (P03)

Building Researcher Mindset through Presentation and Research Encounters Beyond project work, presentation activities, like paper talks, pitches, poster sessions, and mock conferences, helped students clarify ideas, justify design choices, and see themselves as researchers. For many, explaining their work aloud deepened their own understanding. These public moments shifted their mindset from "student" to "researcher." *"The poster session made it feel real. I had to explain our design choices to actual people."* (P07)

Alongside these in-class events, guest lectures and PhD panels exposed students to an authentic research culture and diverse career trajectories. Some drew project ideas from guest talks; others found inspiration in hearing early-stage researchers describe their struggles and decisions. *"Guest lectures… such a good insight into what cutting edge research looks like."* (P02) These exposure events helped students see that research was not an abstract goal or a mysterious domain, but a concrete activity done by real people at different stages, *"Before this class, I thought research was just for grad students or TAs. I didn't know it was something I could understand or even do."* (P10)

Together, these communication and exposure activities helped students develop not only research skills but a sense of identity, audience, and belonging in the research community.

What Challenges Did Students Encounter, and Where Did the Scaffolding Fall Short?

While students overall reported strong gains, they identified a key challenge: balancing structure and autonomy to sup-

Scaffold	Targeted Function	Use Moment	How it operates	Student Challenges & Suggestions
Instructor-led				
Guided in-class activities	Provide conceptual and methodological foundations of ML research	Early course foundational modules	Instructor-guided paper reading, method walk-throughs, and other in-class activities	Technical content inaccessible for some with less technical background; Provide self-assessment modules for background appropriate project definitions..
Milestone check-ins	Provide timely feedback and scope calibration on projects	Proposal; mid-replication; early extension	20-minute project check-ins with the instructor to assess progress	Some check-ins too brief; unclear "what counts" as a valid extension; More structured check-ins; provide early assessment on proposed extensions.
Material-based				
Reading, writing, meeting templates	Guide independent research activities with structured artifacts	After assigned readings and weekly group meetings	Individual summaries → group synthesis; meeting minutes; conference-style template + mock submission	Early readings felt disconnected from later project; Align readings with project areas and pull project scoping earlier.
Reproducibility checklist, setup guides	Lower barrier by structuring replication workflow	Replication warm-up	Provide stepwise checklist to help students setup replications	Buggy repos caused delayed progress or topic changes; Incentivize using a detailed reproducibility checklist.
Peer-based				
Peer feedback rounds	Develop critical thinking skills and provide peer support	After reading presentations, in-group reviews, and final project papers and presentations	Solicit both positive and constructive feedback; mock-up conference to build research identities	Variable feedback quality; Assessments on peer-feedback.

Table 1: Scaffolding Strategies: What Each Support Does, When It is Used, How It Operates, and What are Students' Feedback. The final column color-codes students' proposed issues and adjustments.

port diverse learners.

Insufficient Support for Technical Challenges Students with less technical course prerequisites struggled with technical depth, particularly in understanding ML models: *"I wasn't sure I actually understood what the model did."* (P08) While project phases like replication and extension fostered authentic engagement through increased autonomy, students noted the need for more targeted support in open-ended tasks. For example, technical issues in the replication phase—such as broken or undocumented code—could pose a problem: *"We ended up switching because the original setup was too hard to work with."* (P04)

Module Integration Challenges While students recognized the value of structured paper summary formats in building research reading skills, some found them disconnected from their later project work: *"Now you're going to do a lit review for the project that you've chosen... I feel like that would be a little bit more valuable."* (P04)

In contrast to the majority sentiment (see below), some of the students found the instructor check-ins too brief and suggested structured check-in prompts, that could help with pacing the meetings. Some students described moments where they had promising ideas for extension but ran out of time to implement them. This was not a lack of ideas, but a breakdown in milestone pacing: a reality of exploratory research. This points to the need for additional scaffolds for managing scope, ambition, and the time-bounded nature of a semester-long experience.

These reflections point to a broader insight: scaffolded activities must be well-designed individually, and integrated across time and task boundaries, and this integration must be explicitly communicated to students potentially through additional assignments. Table 1 further summarizes the challenges students faced with different scaffolds, as well as their suggested improvements.

How Do Course Support Structures Enable Scalable Research Learning Across Diverse Backgrounds?

In a large-scale classroom environment where one-on-one mentorship is infeasible, the MLRE course implemented a scaffold composed of instructor check-ins, targeted TA support, and peer feedback mechanisms. While these supports did not eliminate all challenges, they provided anchoring points that helped students stay on track and made it possible for a single instructor to support a large and diverse student population.

Instructor Feedback In the MLRE course, instructor feedback was primarily delivered through periodic group check-ins. Across interviews, students repeatedly emphasized that instructor check-ins played a significant motivational and organizational role: *"She hears all your specific questions that you have, tries to help you like, like frame it, and we keep going. Those check-ins were like the best* *part of the course."* (P06) These check-ins provided effective and pacing support—acting as lightweight scaffolds that helped students feel seen and guided in a large, distributed classroom environment. *"We really just planned around the check-ins."* (P03)

Peer Feedback One of the course's most scalable support mechanisms came through student-to-student learning, particularly within team-based projects and during public presentation moments. Many students appreciated the opportunity to divide labor, engage in shared ownership, and learn from peers' complementary strengths: *"We've got to work on different parts, some of our group members are good at technical implementation, and I can learn from them.*

Meanwhile, public presentations—such as the poster session, which served not just as communication milestones, but also as scaffolds for reflection: *"Hearing how other people described their problems gave us ideas for how to explain ours." (P01)* And presenting to external audiences gave students a stronger sense of ownership and authenticity: *"The poster session made it feel real. I had to explain our design choices to actual people." (P07)*

How Did Students Experience Growth Through the Course?

Students also demonstrated growth in research identity, confidence, and transferable skills, extending beyond the classroom.

From Learners to Researchers: Identity Through Participation MLRE supported students' transition from passive learners to emerging researchers. Scaffolded experiences—replicating papers and proposing extensions—allowed students to see themselves as capable contributors: *"I started asking not just what the model does, but why it does this way, and what else could be tried." (P10)*

Poster sessions, check-ins, and writing assignments became occasions for identity affirmation: *"The poster session made it feel real. I had to explain our design choices to actual people." (P07)*

A Research Onramp for Students from Diverse Backgrounds The course created access points for students with limited or unrelated prior research experience: *"Before this class, I thought research was just for grad students or TAs. I didn't know it was something I could understand or even do… now I feel more confident..." (P10)*

Some used the course to pivot fields or explore long-term goals: *"The course inspired me to do more research.... I took it to get prepared to be a good PhD applicant." (P04)*

Research Capacities as Broadly Transferable Literacy Even students not pursuing research careers found lasting value in the practices taught—reading papers, explaining ideas, and justifying design decisions. These capacities aligned with goals in industry, internships, and technical communication: *"Reading papers is useful for industry too... you can see what others are doingt."* (P06)

The emphasis on technical explanation and public presentation also resonated with students preparing for job interviews: *"It taught me how to speak about technical things to a broader audience… how to explain what we did, why it matters, and what's next. I think that's helpful even in a job interview." (P08)*

Taken together, these findings affirm the value of structured undergraduate research courses. Even in large-scale classrooms, well-designed scaffolds can support students' identity development, broaden access to research pathways, and equip learners with durable literacy that extends beyond academia.

Discussion and Conclusion

MLRE demonstrates that structured, course-based undergraduate research can be both feasible and impactful at scale. By supporting over 90 CS students through a semester-long scaffolded curriculum without individualized mentorship, MLRE offers a scalable alternative that aligns with the goals of broadening participation and epistemic engagement commonly emphasized in URE and CUREs literature (Auchincloss et al. 2014; Russell, Hancock, and McCullough 2007; Linn et al. 2015).

Students reported increased research self-efficacy, appreciation for computational inquiry, and a growing sense of identity as researchers—echoing findings from prior CUREs work (Russell, Hancock, and McCullough 2007; Bangera and Brownell 2014). Notably, MLRE achieved similar developmental outcomes at scale, suggesting that key mechanisms—scaffolded autonomy, modular design, and iterative practice—can effectively support CS research education in large-enrollment settings. Building on these observations, we propose four replicable design principles and areas for future refinement:

1. **Modular Curriculum and Project Alignment.** MLRE structured both lectures and assignments around distinct research phases—literature review, replication, and extension—creating a coherent project workflow. Such modularity reflects characteristics of effective CUREs: iteration, discovery-oriented tasks, and relevance to external communities (Auchincloss et al. 2014).
2. **Scaffolded Progression from Replication to Autonomy.** By framing tasks to move gradually from well-defined replication to student-designed extensions, MLRE operationalized the principle of scaffold fading (Wood, Bruner, and Ross 1976). This progression lowered entry barriers while fostering genuine inquiry.
3. **Differentiated Support for Diverse Learners.** To accommodate heterogeneous backgrounds, the course included milestone check-ins, rubrics tuned to task complexity, and both TA and peer feedback. Future iterations will explore more adaptive supports—peer mentoring lanes, early diagnostics, and automated checkpoints—to sustain engagement across readiness levels.

Altogether, these strategies instantiate a transferable model for embedding research into the undergraduate CS curriculum.We hope this work contributes to ongoing efforts to make undergraduate research a routine, equitable part of the CS curriculum—and invites broader experimentation with research-based pedagogies at scale.

Acknowledgments

We thank the students and teaching assistants in the Machine Learning Research Experience course for their contributions and feedback throughout this course; especially those who participated in the focus groups. This work was supported in part by the Inclusive-Teaching Professional Development Fund and the Departmental Summer Course Development Fund. We also appreciate the helpful comments from Tershia Pinder-Grover as well as her role in soliciting student feedback and ultimately improving student experience in the course.

References

Alvarado, C.; Villazon, S.; and Tamer, B. 2019. Evaluating a scalable program for undergraduate CS research. In *Proceedings of the 2019 ACM Conference on International Computing Education Research*, 269–277.

Ambrose, S. A.; Bridges, M. W.; DiPietro, M.; Lovett, M. C.; and Norman, M. K. 2010. *How learning works: Seven research-based principles for smart teaching*. John Wiley & Sons.

Anders Ericsson, K. 2008. Deliberate practice and acquisition of expert performance: a general overview. *Academic emergency medicine*, 15(11): 988–994.

Auchincloss, L. C.; Laursen, S. L.; Branchaw, J. L.; Eagan, K.; Graham, M.; Hanauer, D. I.; Lawrie, G.; McLinn, C. M.; Pelaez, N.; Rowland, S.; Towns, M.; Trautmann, N. M.; Varma-Nelson, P.; Weston, T. J.; and Dolan, E. L. 2014. Assessment of Course-Based Undergraduate Research Experiences: A Meeting Report. *CBE—Life Sciences Education*, 13(1): 29–40.

Bangera, G.; and Brownell, S. E. 2014. Course-based undergraduate research experiences can make scientific research more inclusive. *CBE—Life Sciences Education*, 13(4): 602–606.

Bekkering, E. 2025. Course-Based Undergraduate Research Experiences (CUREs) for Computer Science?. *Information Systems Education Journal*, 23(3): 4–15.

Belland, B. R.; and Belland, B. R. 2017. Instructional scaffolding: Foundations and evolving definition. *Instructional Scaffolding in STEM Education: Strategies and efficacy evidence*, 17–53.

Carrabba, C.; and Farmer, A. 2018. The impact of project-based learning and direct instruction on the motivation and engagement of middle school students. *Language Teaching and Educational Research*, 1(2): 163–174.

Chi, M. T.; and Wylie, R. 2014. The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, 49(4): 219–243.

Connors, P. K.; Lanier, H. C.; Erb, L. P.; Varner, J.; Dizney, L.; Flaherty, E. A.; Duggan, J. M.; Yahnke, C. J.; and Hanson, J. D. 2021. Connected while distant: networking CUREs across classrooms to create community and empower students. *Integrative and Comparative Biology*, 61(3): 934–943.

Dougherty, R. E. 2024. Scaffolding Research Projects in Theory of Computing Courses. *arXiv preprint arXiv:2410.01712*.

Fernandez, A. S. 2025. A Course-based Undergraduate Research Experience (CURE) Focused Broadly on Research Methods in Computer Science. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1*, 332–338.

Fund, F. 2023. We need more reproducibility content across the computer science curriculum. In *Proceedings of the 2023 ACM Conference on Reproducibility and Replicability*, 97–101.

Goodlad, S. 1998. Research opportunities for undergraduates. *Studies in Higher Education*, 23(3): 349–356.

Granger, M. J.; Amoussou, G.-A.; Labrador, M. A.; Perry, S.; and Van Busum, K. M. 2006. Research experience for undergraduates: successes and challenges. *ACM SIGCSE Bulletin*, 38(1): 558–559.

Hutson, M. 2018. Artificial intelligence faces reproducibility crisis. *Science*, 359(6377): 725–726.

Klahr, D.; and Nigam, M. 2004. The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological science*, 15(10): 661–667.

Kutty, S.; and Guzdial, M. 2020. Undergraduate Student Research With Low Faculty Cost. volume 1, 1–2. IEEE.

Linn, M. C.; Palmer, E.; Baranger, A.; Gerard, E.; and Stone, E. 2015. Undergraduate research experiences: Impacts and opportunities. *Science*, 347(6222): 1261757.

Lu, X.; Fan, S.; Houghton, J.; Wang, L.; and Wang, X. 2023. ReadingQuizMaker: a human-NLP collaborative system that supports instructors to design high-quality reading quiz questions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–18.

Lucic, A.; Bleeker, M.; Jullien, S.; Bhargav, S.; and De Rijke, M. 2022. Reproducibility as a mechanism for teaching fairness, accountability, confidentiality, and transparency in artificial intelligence. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 12792–12800.

Martin, N. D.; Dornfeld Tissenbaum, C.; Gnesdilow, D.; and Puntambekar, S. 2019. Fading distributed scaffolds: The importance of complementarity between teacher and material scaffolds. *Instructional Science*, 47: 69–98.

McNeill, K. L.; Lizotte, D. J.; Krajcik, J.; and Marx, R. W. 2006. Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The journal of the Learning Sciences*, 15(2): 153–191.

Pineau, J. e. a. 2020. Improving reproducibility in machine learning research (A report from the NeurIPS 2019 reproducibility program). *arXiv preprint arXiv:2003.12206*.

Russell, S. H.; Hancock, M. P.; and McCullough, J. 2007. Benefits of Undergraduate Research Experiences. *Science*, 316(5824): 548–549.

Stephens, A.; Brenner, K.; and Gentile, J. 2017. Undergraduate research experiences for STEM students: Successes, challenges, and opportunities.

Vygotsky, L. S.; and Cole, M. 1978. *Mind in society: Development of higher psychological processes*. Harvard university press.

Werner, L.; Layaïda, N.; Genevès, P.; Euzenat, J.; and Graux, D. 2024. Reproduce, replicate, reevaluate. the long but safe way to extend machine learning methods. In *Proceedings of* *the AAAI Conference on Artificial Intelligence*, volume 38, 15850–15858.

Wood, D. J.; Bruner, J. S.; and Ross, G. 1976. The Role of Tutoring in Problem Solving. In *Journal of Child Psychology and Psychiatry*, 89–100. Wiley.

Yildiz, B.; Hung, H.; Krijthe, J. H.; Liem, C. C.; Loog, M.; Migut, G.; Oliehoek, F. A.; Panichella, A.; Pawełczak, P.; Picek, S.; et al. 2021. ReproducedPapers. org: Openly teaching and structuring machine learning reproducibility. In *International Workshop on Reproducible Research in Pattern Recognition*, 3–11. Springer.