STEM Kits as Portable Practice-Based Learning Environments: A Narrative Literature Review With Implications for Advanced Manufacturing

A Granite Hearth White Paper
Prepared by Richard Feistman, Ph.D., Executive Director
September 2026

Suggested Citation:
Granite Hearth. (2026). *STEM kits as portable practice-based learning environments: A narrative literature review with implications for advanced manufacturing*.

Abstract

This narrative review synthesizes research on active learning, guided inquiry, concrete manipulatives, maker education, engineering design, informal science learning, and advanced manufacturing education to clarify how STEM kits may support learning. The evidence does not justify treating hands-on materials as effective interventions by themselves. Rather, physical kits appear most defensible as instructional mediators that can make selected concepts or disciplinary practices observable and actionable when they are paired with explicit learning goals, appropriate guidance, opportunities for reflection, and authentic application. Direct studies of electronics, additive-manufacturing, invention, and fluid-power kits provide promising examples but a comparatively limited causal evidence base. Research on maker education and learning factories similarly suggests potential for engagement, competence development, and self-efficacy while highlighting uneven assessment quality and implementation challenges. This review proposes the term portable practice-based learning environment as a working design concept for kits that approximate selected features of scientific, engineering, manufacturing, or research practice. MBADIKA's M°Lab Field Kit 2.0 and Granite Hearth's Four Stories Qualitative Investigator Kit are examined as practice-based examples. The review concludes with design and evaluation principles for future Granite Hearth kits.

Keywords: STEM kits, maker education, guided inquiry, advanced manufacturing, learning factories, practice-based learning

STEM Kits as Portable Practice-Based Learning Environments: A Narrative Literature Review With Implications for Advanced Manufacturing

Review Question and Scope

This review asks: How and under what conditions do STEM kits support conceptual learning, disciplinary practice, engagement, identity, and access, particularly when they approximate authentic scientific, engineering, or manufacturing work?

The review is narrative rather than systematic. It combines sources assembled in Granite Hearth's existing work on STEM kits with targeted additions from peer-reviewed systematic reviews, meta-analyses, engineering-education studies, National Academies syntheses, federal reports, and selected practice-based materials. The evidence base is intentionally broader than studies that use the exact term STEM kit because kit-mediated learning intersects with several research traditions, including active learning, inquiry, maker education, manipulatives, engineering design, informal science learning, and learning factories. Consequently, broad findings from adjacent literatures are treated as explanatory context rather than as direct evidence that a particular kit is effective.

From Hands-On to Practice-Based Learning

Contemporary STEM education increasingly emphasizes participation in disciplinary practices rather than recall of content alone. The National Research Council (2012) identified eight science and engineering practices, including asking questions and defining problems, developing and using models, planning and carrying out investigations, analyzing and interpreting data, constructing explanations and designing solutions, arguing from evidence, and communicating information. Integrated STEM frameworks likewise emphasize connections among disciplines and engagement with authentic problems rather than isolated subject instruction (National Academy of Engineering & National Research Council, 2014; Thibaut et al., 2018).

This broader movement toward practice-oriented STEM learning provides the intellectual context for STEM kits. A kit is not valuable simply because it contains physical objects. Its educational value depends on whether the materials and tasks organize meaningful participation in scientific or engineering activity. A learner who follows a recipe without understanding may be physically active but conceptually passive; a learner who tests a material, interprets a result, revises a design, and connects the result to a disciplinary principle is engaged in more substantive STEM practice.

The broader active-learning literature is often invoked to justify hands-on approaches, but it should be used carefully. Freeman et al. (2014) synthesized 225 undergraduate STEM studies and found higher examination and concept-inventory performance and lower failure under active learning than under traditional lecture. That finding strongly supports active participation as an instructional principle, but it is not direct evidence that kits themselves cause learning. The more precise question is which features of kit-mediated activity generate useful cognitive, social, and disciplinary work.

What Is a STEM Kit?

For purposes of this review, a STEM kit is a structured set of physical materials, components, representations, and instructional supports designed to enable learners to investigate, build, test, model, measure, or analyze STEM-related ideas or practices. This definition is intentionally broader than a box of experiment supplies. It includes modular electronics systems, manufacturing models, materials-testing kits, invention kits, robotics or fabrication tools, and other bounded physical environments that organize learner activity.

This definition also distinguishes a kit from curriculum and pedagogy. Curriculum specifies learning goals, content, sequence, and expected outcomes. Pedagogy specifies how learning is facilitated. A kit operates primarily at the level of interaction: it changes what learners can see, manipulate, measure, compare, and produce. The same kit can therefore generate different outcomes depending on curriculum, facilitation, time available, and learner context.

As a working definition proposed in this review, a STEM kit can be understood as a portable practice-based learning environment: a deliberately bounded collection of tools, materials, representations, and prompts that gives novices access to selected concepts and disciplinary practices that might otherwise require specialized equipment, expertise, or facilities.

Theoretical Foundations

Active Learning and Guided Inquiry

Active learning provides a broad foundation for kit-mediated instruction. The Freeman et al. (2014) meta-analysis found a robust advantage for active learning across undergraduate STEM disciplines. However, active learning is a family of approaches rather than a single intervention, and the effectiveness of active participation depends on how the activity is structured.

Guided inquiry is particularly relevant. Lazonder and Harmsen (2016) synthesized 72 studies and found facilitative effects of instructional guidance on inquiry activities, successful task performance, and learning outcomes. Their findings challenge the assumption that giving learners materials and freedom to explore is sufficient. Productive inquiry generally benefits from prompts, explanations, heuristics, scaffolds, feedback, or task constraints. For kit design, instructions and facilitation are therefore part of the learning mechanism rather than peripheral add-ons.

Concrete Representations and Concreteness Fading

A second foundation comes from research on concrete manipulatives. Carbonneau et al. (2013) synthesized 55 mathematics studies involving 7,237 learners from kindergarten through college and found small-to-moderate advantages for instruction using concrete manipulatives over instruction using abstract symbols alone. The magnitude of benefits varied by learning outcome and by instructional and methodological features. Because this evidence is from mathematics instruction rather than STEM kits specifically, it is best treated as adjacent evidence about the potential role of tangible representations.

Fyfe et al. (2014) offer a useful design principle for addressing the limits of concrete representations. Their systematic review of concreteness fading recommends beginning with concrete materials and then explicitly and gradually linking them to more abstract representations. Applied to STEM kits, the implication is that the physical object should function as a bridge rather than an endpoint: learners should move from manipulating a component, to observing a relationship, to naming the underlying concept, and then to applying that concept beyond the original kit.

Situated Cognition and Authentic Practice

Situated cognition adds another layer. Brown et al. (1989) argued that knowledge is situated and is partly a product of the activity, context, and culture in which it is developed and used. This perspective supports learning environments that do more than explain what professionals know; they approximate what professionals do. In STEM education, this is consistent with the National Research Council's (2012) emphasis on student participation in science and engineering practices derived from the work of scientists and engineers.

For kits, situated learning suggests that authenticity can matter, although authenticity should not be equated with perfect replication of professional equipment. A learner can study fabric properties as a list of terms, for example, or investigate fabric performance while making a product decision. The latter creates a reason for the knowledge and places concepts within a recognizable practice. In a portable kit, authenticity may therefore be approximated through selected decisions, evidence, tools, constraints, and disciplinary language.

Engineering Design and Integrated STEM

Integrated STEM research reinforces the value of problem-centered, inquiry-based, design-based, and collaborative learning. Thibaut et al. (2018) identified integration of STEM content, problem-centered learning, inquiry-based learning, design-based learning, and cooperative learning as recurring instructional principles in secondary integrated STEM education. Le et al. (2023), reviewing 47 K-12 studies, likewise found that outcomes differed across integrated STEM approaches and educational settings. The implication for kits is outcome specificity: there is no single STEM effect, so activity structure and intended outcomes need to be aligned.

Crismond and Adams (2012) provide a useful engineering-design lens through the Informed Design Teaching and Learning Matrix. The framework contrasts beginning and more informed designers across practices that include understanding the challenge, building knowledge, generating ideas, experimenting, troubleshooting, iterating, and reflecting. Kits that deliberately create opportunities for these practices more closely approximate engineering design than kits that culminate only in a predetermined build.

What the Maker-Education Literature Adds

Maker education is the closest large research tradition to many contemporary STEM kits because it combines physical fabrication, digital tools, design, and artifact production. The evidence is encouraging but nuanced. Davies and Seitamaa-Hakkarainen (2025) conducted a systematic review of 62 K-12 maker-education studies that met the authors' inclusion and quality criteria. The reviewed studies reported a range of learning, design, identity, and community-building outcomes, and many combined tangible materials with digital tools or fabrication technologies. At the same time, the review highlighted two studies showing that science and mathematics conceptual knowledge cannot be assumed to emerge naturally from making; teachers need to identify, design, and focus tasks on the disciplinary concepts they intend students to learn.

A 2026 meta-systematic review by Elmoazen et al. synthesized 17 systematic and scoping reviews of makerspace education published from 2019 through 2024. Across those reviews, commonly reported benefits included creativity, 21st-century skills, academic performance, motivation, social responsiveness, and digital literacy. The synthesis also identified persistent challenges involving curriculum integration, educator training, and equitable access. These findings reinforce the central claim of this review: the learning environment around a kit matters as much as the presence of the kit itself.

Assessment quality remains a limitation in this literature. Lin et al. (2020) reviewed 60 empirical studies of technology-rich maker activities and found more than 20 types of maker platforms and five prevalent assessment formats: artifact assessments, tests, surveys, interviews, and observations. Only 15% of the studies reported psychometric evidence for the reliability or validity of their assessments. Positive reports about engagement or learning should therefore be interpreted in light of a still-developing measurement base.

Direct Evidence on Kits and Physical Learning Tools

Within the literature identified for this narrative review, direct empirical research on bounded STEM kits is less extensive than the adjacent active-learning, inquiry, and maker literatures. The available studies are nevertheless useful because they illustrate both the promise and the limitations of kit-mediated learning. Sabogal Guerrero (2023) developed a modular electronics educational kit using magnetically connectable blocks to support learning about the Wheatstone bridge, signal filtering, and transistor amplification. The project emphasized independent learning, logical reasoning, and critical thinking. As an undergraduate thesis centered on design and development, it provides evidence about a kit concept and proof of concept rather than strong causal evidence about learning outcomes. Valjak et al. (2023) developed an additive-manufacturing educational kit consisting of 17 physical models and accompanying cards with short explanations, graphics, examples, and manufacturing data. The authors positioned the kit as an introductory tool for novices learning basic characteristics and terminology of material-extrusion additive manufacturing before advancing to more complex additive-manufacturing and design-for-additive-manufacturing topics. The study is a development paper rather than a controlled effectiveness evaluation. Metwaly et al. (2025) developed a fluid-power kit with five modules for engineering technology students. Learners constructed grippers and basic deadweight testers, used Arduino-based data acquisition, calibrated sensors, and visualized data on LCD screens. The authors reported positive changes in students' understanding of the material, engineering identity, and ability to apply theoretical knowledge to manufacturing applications. Because the evidence comes from a conference study using course-based surveys and instructional data rather than a randomized design, these results should be treated as promising rather than definitive.

A particularly important cautionary example is the external evaluation by LiCalsi et al. (2019), which examined a middle-school physical science and engineering invention-kit curriculum with 419 students in grades 7 and 8. The quasi-experimental evaluation found no statistically significant differences in physical science and engineering knowledge or in STEM interest and confidence between students using the kits and comparison students. Implementation fidelity was also uneven: the study met its agreed fidelity standard on only one of three key components. The What Works Clearinghouse subsequently rated the study as meeting standards with reservations and reported no statistically significant positive findings (What Works Clearinghouse, 2021). This example is important because it prevents an overly promotional conclusion: well-developed materials do not guarantee measurable effects when implementation or alignment is weak.

Learning Outcomes That Kits May Support

Conceptual Understanding

The most defensible conceptual-learning mechanism is the translation of abstract relationships into observable systems. Adjacent evidence from mathematics suggests that concrete representation can improve learning relative to abstract-only instruction (Carbonneau et al., 2013), while concreteness-fading research emphasizes the importance of explicitly linking concrete experiences to generalized concepts (Fyfe et al., 2014). Direct kit studies in electronics, fluid power, and additive manufacturing are consistent with this mechanism because they use physical systems to make technical relationships visible, testable, or discussable (Metwaly et al., 2025; Sabogal Guerrero, 2023; Valjak et al., 2023).

The limitation is equally important. Maker research indicates that conceptual learning does not necessarily follow from artifact production alone (Davies & Seitamaa-Hakkarainen, 2025). A learner may successfully build an object through imitation, trial and error, or peer assistance without forming a transferable understanding of the relevant science or mathematics. Effective kit design therefore requires an explicit bridge between doing and explaining.

Engagement, Motivation, Self-Efficacy, and Identity

Hands-on and maker environments frequently report outcomes related to motivation, engagement, confidence, and self-efficacy. These affective outcomes are relevant because STEM participation is influenced in part by learners' beliefs about their capabilities and by whether they can imagine themselves participating in a field. Okundaye et al. (2022) provide a particularly relevant manufacturing example through their Making as Micro-Manufacturing model, which integrated fabrication with real production requirements for high-school students from an underserved community. The authors reported descriptive increases across dimensions of making and engineering self-efficacy during participation in a real production process. Because the implementation involved a very small group and relied heavily on self-report, the findings are best treated as suggestive rather than generalizable evidence.

The same study also shows why authentic practice should not be reduced to a fun-equals-learning narrative. During the production cycle, self-reported making and engineering self-efficacy varied as students worked under real-world deadlines and expectations (Okundaye et al., 2022). Authentic production can introduce challenge, constraint, quality expectations, and iteration. Those demands may be educationally valuable even when they do not make the experience uniformly easy or enjoyable.

Problem Solving, Collaboration, and Career Awareness

Integrated STEM, maker education, and engineering-design research place problem solving and collaboration near the center of learning. Making and design projects can require learners to negotiate roles, troubleshoot failures, seek help, combine knowledge, and revise artifacts (Crismond & Adams, 2012; Davies & Seitamaa-Hakkarainen, 2025). These practices are relevant to technical work because they require learners to act on evidence and coordinate decisions rather than complete isolated procedural tasks.

Career awareness is a related but distinct outcome. The Learn&Fly STEM Kit, for example, was designed to foster engagement in science subjects and aeronautics-related careers through glider design and construction, supported by student, teacher, and career resources (Guedes et al., 2019). As a practice-oriented instructional resource rather than a strong causal evaluation, it is most useful here as a design example: the activity is embedded in a recognizable occupational and technical context. A similar principle may be valuable for advanced manufacturing kits intended to expose learners to materials, sensors, fabrication, testing, or production decision-making.

Conditions for Effectiveness

Guidance is one of the clearest recurring conditions across the evidence. Inquiry-learning research shows that guidance improves inquiry activity, successful performance, and learning outcomes (Lazonder & Harmsen, 2016). Maker-education research likewise highlights embedded scaffolding, reflective discussion, and facilitator or peer support (Davies & Seitamaa-Hakkarainen, 2025). A kit designed for novice use should therefore include more than assembly directions; it should support prediction, observation, comparison, explanation, and reflection.

The physical task should be chosen because it reveals a target concept or practice, not simply because it is engaging. Davies and Seitamaa-Hakkarainen (2025) show why this matters: maker activity does not guarantee science or mathematics learning unless those concepts are intentionally foregrounded. Kit design should therefore specify what learners are expected to notice, what evidence they should collect, and how the activity connects to a disciplinary idea.

Implementation conditions also matter. The invention-kit evaluation by LiCalsi et al. (2019) found no statistically significant effects and documented that only one of three agreed fidelity components was met across the study as a whole. Early-STEM teacher research identifies related practical constraints involving time, support, resources, teacher capacity, and subject knowledge (Wan et al., 2021). For community-based kit programs, these findings support the use of facilitator guides, concise training, troubleshooting support, and realistic estimates of setup and instructional time.

Kits are often proposed as access tools because they can potentially bring selected learning experiences into schools, libraries, after-school programs, homes, or community organizations without full laboratory infrastructure. That is a plausible design advantage, not a demonstrated universal equity effect. Makerspace reviews identify continuing challenges involving curriculum integration, educator preparation, and equitable access (Elmoazen et al., 2026). The National Research Council (2009) also emphasizes that science learning occurs across everyday, designed, and community settings. The equity value of portable kits is therefore best framed as the potential to reduce selected infrastructure and participation barriers while still requiring attention to language, prior knowledge, disability access, facilitator capacity, cultural relevance, and meaningful learner choice.

The literature is more mature in describing promising experiences than in demonstrating durable causal effects. Many studies rely on small samples, self-report, short interventions, or locally developed measures. Lin et al. (2020) documented limited psychometric reporting in maker assessment. Reining and Kauffeld (2022) similarly found that empirical evaluations of learning factories varied substantially in design and rigor, with only two of 22 underlying studies using control groups. These limitations have direct implications for Granite Hearth: evaluation should be built into kit implementation rather than treated as a later add-on.

Advanced Manufacturing: From Learning Factories to Portable Kits

Learning Factories as a High-Fidelity Comparison

Advanced manufacturing education provides a useful comparison because the field has invested in practice-oriented training environments. Abele et al. (2017) describe learning factories as environments for education, training, and research in manufacturing and review their didactic foundations, curricula, applications, potential, and limitations. In broad terms, learning factories model important features of production processes and value chains so that learners can develop competence through practice rather than through explanation alone.

The empirical learning-factory literature is promising but methodologically uneven. Reining and Kauffeld (2022) identified 24 empirical publications representing 22 underlying studies of learning success or competence development at learning factories. The reviewed studies reported improvements across professional, methodological, social, and self-competence domains, but much of the evidence relied on self-assessment, and only two studies used control groups. Roll and Ifenthaler (2021), using three competence tests with 63 technical vocational students, found evidence of improvement in subject-related technical competencies but did not identify positive development in broader multidisciplinary digital competencies. These results reinforce the need to specify the competency a learning environment is intended to develop.

Portable Kits as Partial Approximations of Practice

This review proposes that portable kits can be viewed as a lower-fidelity, lower-infrastructure point on the same practice-based continuum as learning factories. This is a conceptual synthesis rather than a direct empirical finding from the learning-factory literature. A portable kit cannot reproduce a production line or industrial laboratory, and it should not be marketed as doing so. It can, however, extract a coherent subset of professional practice and make that subset available in settings without specialized facilities. A manufacturing kit might expose learners to materials, measurement, sensors, process variation, data interpretation, quality constraints, design trade-offs, or iterative testing.

This formulation also clarifies what authenticity can mean in a portable format. Authenticity need not require full workplace simulation. A bounded experience can approximate practice when learners confront a meaningful problem, use representative tools or evidence, make decisions using disciplinary criteria, and explain the consequences of those decisions (Brown et al., 1989; National Research Council, 2012). The design goal is to preserve selected intellectual and procedural features of the practice even when equipment has been simplified.

Making as Micro-Manufacturing

Okundaye et al. (2022) provide one of the clearest bridges between maker education and manufacturing. Their Making as Micro-Manufacturing model couples fabrication practices with production concerns and real-world product requirements. In the reported implementation, high-school students produced small-batch products, including instructional science kits for elementary-school learners. The model is especially relevant to Granite Hearth because it suggests that the educational value of kits can extend beyond the person who receives a kit. Designing, assembling, testing, documenting, packaging, or improving kits can itself become a manufacturing-learning experience. In other words, kits can function both as learning products and as objects through which students learn production.

MBADIKA's M°Lab Field Kit 2.0

MBADIKA's M°Lab Field Kit 2.0 illustrates how a portable kit can introduce materials science and product redesign through an advanced-fabric context. Developed through a collaboration between MBADIKA and Ministry of Supply, the kit places learners in the role of an M°Lab Secret Agent charged with redesigning a uniform. Learners explore fabric technology and construction, conduct performance tests, and examine the chemistry of dyeing natural and synthetic fabrics (MBADIKA, n.d.). The narrative provides a reason for the experiments: material properties matter because they inform a product decision.

Figure 1
MBADIKA M°Lab Field Kit 2.0

Note. Image courtesy of MBADIKA. Reproduced with permission.

The kit is also connected to the advanced-manufacturing workforce ecosystem. In the Manufacturing USA 2025 Annual Report, Jones et al. (2026) describe an AFFOA partnership with Mbadika and Ministry of Supply to scale up the M-Lab Field Kit 2.0 for middle- and high-school students. The report states that Mbadika hosted 21 workshops and a textile-education session that engaged more than 700 participants across Massachusetts, Rhode Island, and Pennsylvania. Those figures demonstrate reach and implementation activity, not causal learning effects.

M°Lab is therefore best treated in this review as an evidence-informed design example rather than as an empirically validated intervention. Its design reflects several characteristics supported by adjacent literature: authentic framing, tangible materials, guided investigation, application to a product decision, and exposure to a contemporary manufacturing domain. A future evaluation could test outcomes such as materials-science understanding, knowledge of advanced-textiles careers, design reasoning, or manufacturing self-efficacy.

Implications for Granite Hearth and the Four Stories Kit

Granite Hearth's Four Stories Qualitative Investigator Kit is not an advanced manufacturing kit and should not be represented as one. Its relevance is conceptual: it

Figure 2
4 Stories, Granite Hearth

demonstrates the same portable practice-based design logic in a different research domain. The kit uses question cards, follow-up prompts, note cards, theme cards, a wet-erase marker, and an instructional guide to move a learner through interviewing, note-taking, qualitative coding, theme development, interpretation, and theoretical memo writing (Granite Hearth, 2026).

Within the design framework proposed in this review, the physical cards externalize parts of a complex analytic process so that a novice can see and manipulate steps that experienced researchers may perform mentally or with specialized software. Note cards function as units of data; theme cards function as provisional analytic categories; and physically sorting notes by theme makes the logic of categorization visible. This is an interpretation of the kit design rather than an empirical finding about its effectiveness.

Table 1

Two Examples of Portable Practice-Based Kit Design

Design dimension	M°Lab Field Kit 2.0	Four Stories Qualitative Investigator Kit
Role or scenario	M°Lab agent in a uniform-redesign scenario	Qualitative researcher
Tangible tools	Fabric samples and testing materials	Question, prompt, note, and theme cards
Core activity	Investigate fabric properties and interpret performance	Collect stories, code evidence, and develop themes
Evidence used	Observed material behavior	Interview notes and patterns across accounts
Decision or product	Redesign a uniform and justify material choices	Construct an evidence-based interpretation and theoretical memo
Practice approximated	Materials testing, product redesign, and design trade-offs	Interviewing, qualitative analysis, and analytic memo writing
Primary opportunity	Connect materials science to advanced fabrics and manufacturing	Make an abstract research process visible and performable

Note. M°Lab details are based on MBADIKA (n.d.); Four Stories details are based on Granite Hearth (2026). Labels describing the practices approximated are interpretive synthesis for this review.

The comparison suggests a useful organizing principle for Granite Hearth's future kit portfolio: kits can make either phenomena or practices tangible. M°Lab makes material properties and product-design trade-offs tangible. Four Stories makes data collection and qualitative analysis tangible. Future advanced-manufacturing kits could apply the same design logic to sensing, quality control, digital fabrication, process optimization, robotics, composites, or supply-chain decisions.

A Proposed Design Model for Practice-Based Kits

Synthesizing the literature suggests a five-part causal chain. This is a proposed conceptual model, not a tested causal model. It avoids attributing learning directly to the physical kit and instead identifies intermediate mechanisms that a kit is intended to activate.

Table 2

Proposed Mechanism for Kit-Mediated Learning

Kit design	Learner activity	Learning mechanism	Proximal outcomes	Longer-term outcomes
Materials, representations, prompts, constraints	Build, test, compare, interview, measure, analyze	Attention, feedback, externalization, iteration, conceptual linking	Knowledge, confidence, engagement, disciplinary practice	Identity, persistence, career awareness, transfer

Note. Proposed synthesis informed by research on active learning, guided inquiry, concrete-to-abstract representation, maker education, and disciplinary practice (Brown et al., 1989; Davies & Seitamaa-Hakkarainen, 2025; Freeman et al., 2014; Fyfe et al., 2014; Lazonder & Harmsen, 2016). Important moderating conditions include guidance and scaffolding, curriculum alignment, facilitator preparation, implementation fidelity, time on task, prior knowledge, accessibility, cultural relevance, opportunities for reflection, and measurement quality.

Design Principles for Granite Hearth Kits

The following principles are a synthesis of the reviewed evidence rather than a validated checklist. They translate recurring findings about guidance, authentic practice, concrete representation, design activity, and evaluation into development criteria for Granite Hearth kits.

- Begin with a real practice, not with the contents of the box. Define what a scientist, engineer, technician, designer, or researcher actually does and identify a bounded practice that can be performed safely and affordably (Brown et al., 1989; National Research Council, 2012).
- Make an important relationship visible. Use the physical components to help learners observe a relationship that might otherwise remain abstract, while recognizing that concrete materials are most useful when connected to the underlying concept (Carbonneau et al., 2013; Fyfe et al., 2014).
- Build in inquiry, not just assembly. Ask learners to predict, compare, test, interpret, troubleshoot, and revise rather than relying on free exploration or a predetermined build alone (Crismond & Adams, 2012; Lazonder & Harmsen, 2016).
- Connect concrete experience to explicit concepts. Use prompts, diagrams, discussion, or follow-up tasks that help learners name the relevant concept and apply it beyond the immediate object (Davies & Seitamaa-Hakkarainen, 2025; Fyfe et al., 2014).
- Use authentic constraints and decisions. Give learners a reason to care about the evidence, such as choosing a material, diagnosing a failure, meeting a product requirement, interpreting a dataset, improving a design, or communicating a recommendation (Brown et al., 1989; Okundaye et al., 2022).
- Design the facilitator experience as carefully as the learner experience. Provide setup guidance, realistic time estimates, common failure points, discussion prompts, and concise explanations of the concepts the kit is intended to surface (Davies & Seitamaa-Hakkarainen, 2025; Lazonder & Harmsen, 2016; Wan et al., 2021).

- Preserve low barriers to entry without removing productive challenge. Novices need scaffolding, but iteration and troubleshooting are legitimate parts of design practice (Crismond & Adams, 2012).
- Build evaluation into the kit from the start. Define a small set of intended outcomes and use measures that distinguish reach, satisfaction, learning, self-efficacy, identity, and continued participation rather than treating them as interchangeable indicators of success (Lin et al., 2020; Reining & Kauffeld, 2022).

Research Gaps and Evaluation Priorities

The current literature supports continued development and study of STEM kits, but it does not justify broad claims that kits themselves produce learning. The field contains many design examples and positive implementation reports, while fewer studies isolate which features matter, for whom, under what conditions, and with what durability. Granite Hearth can contribute to this evidence base by treating kit development as an evaluation agenda as well as a program strategy.

A practical evaluation framework should distinguish at least five types of outcomes: (a) implementation and reach, such as the number of kits used and completion rates; (b) learner experience, including engagement and perceived relevance; (c) proximal learning, such as concept knowledge or performance of a targeted practice; (d) self-efficacy and identity, such as confidence using tools or seeing oneself participating in a field; and (e) transfer or continued participation, such as later enrollment, additional making, career exploration, or use of the practice in a new context. This separation is important because evidence that a program reached learners or that learners enjoyed it is not evidence of conceptual learning or longer-term persistence (Lin et al., 2020; Reining & Kauffeld, 2022).

For M°Lab or a comparable advanced-manufacturing kit, a small-scale evaluation could combine a brief pre/post concept measure, a performance task requiring learners to justify a material or design decision, a manufacturing-career awareness measure, and facilitator implementation logs. For Four Stories, measures could assess whether learners can distinguish notes from themes, organize evidence into coherent categories, explain the rationale for a theme, and produce a memo whose claims are supported by interview evidence. These measures would test the specific practice each kit is designed to make accessible rather than relying only on satisfaction ratings.

Conclusion

The literature supports a strong but conditional case for STEM kits as tools within well-designed learning environments. Hands-on activity is most educationally meaningful when learners are required to notice, test, explain, decide, and revise. Physical components can make selected relationships visible; guidance can help turn manipulation into reasoning; authentic contexts can give technical knowledge a purpose; and portable formats may reduce some infrastructure barriers. None of these effects is automatic (Davies & Seitamaa-Hakkarainen, 2025; Fyfe et al., 2014; Lazonder & Harmsen, 2016).

The most useful conceptual shift proposed by this review is from the STEM kit as a box of materials to a portable practice-based learning environment. This term is not presented as an established construct in the literature; it is a synthesis intended to connect research on maker education, concrete representations, inquiry, engineering design, informal science learning, and advanced manufacturing learning factories. It also creates a coherent way to understand both MBADIKA's M°Lab Field Kit and Granite Hearth's Four Stories kit. M°Lab makes materials-science phenomena and product decisions tangible, while Four Stories makes a qualitative-research process tangible. In each case, the physical tools organize participation in a bounded practice.

For Granite Hearth, this framing offers a linked design and evaluation strategy: identify a meaningful professional practice, select tools that reveal an important relationship or decision, scaffold concrete action into explicit understanding, and measure the intended outcomes. This is more defensible than a general claim that hands-on learning works.

References

Abele, E., Chryssolouris, G., Sihn, W., Metternich, J., ElMaraghy, H., Seliger, G., Sivard, G., ElMaraghy, W., Hummel, V., Tisch, M., & Seifermann, S. (2017). Learning factories for future oriented research and education in manufacturing. *CIRP Annals - Manufacturing Technology, 66*(2), 803-826. https://doi.org/10.1016/j.cirp.2017.05.005

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42. https://doi.org/10.3102/0013189X018001032

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology, 105*(2), 380-400. https://doi.org/10.1037/a0031084

Crismond, D. P., & Adams, R. S. (2012). The informed design teaching and learning matrix. *Journal of Engineering Education, 101*(4), 738-797. https://doi.org/10.1002/j.2168-9830.2012.tb01127.x

Davies, S., & Seitamaa-Hakkarainen, P. (2025). Research on K-12 maker education in the early 2020s - a systematic literature review. *International Journal of Technology and Design Education, 35*(2), 763-788. https://doi.org/10.1007/s10798-024-09921-6

Elmoazen, R., López-Pernas, S., & Saqr, M. (2026). Mapping makerspaces in education: A meta-systematic review. *International Journal of Technology and Design Education*. Advance online publication. https://doi.org/10.1007/s10798-026-10093-8

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410-8415. https://doi.org/10.1073/pnas.1319030111

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review, 26*(1), 9-25. https://doi.org/10.1007/s10648-014-9249-3

Granite Hearth. (2026). *Four Stories qualitative investigator kit: Instruction insert* [Unpublished instructional material].

Guedes, M., Piteira, M., Nunes, N., & Cláudio, R. (2019). *STEM kit: Teachers' notebook*. Instituto Politécnico de Setúbal. http://hdl.handle.net/10400.26/34913

Jones, C., Tran, M., & Stephens, A. (2026). *Manufacturing USA 2025 annual report* (NIST Advanced Manufacturing Series 600-19). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AMS.600-19

Le, H. C., Nguyen, V. H., & Nguyen, T. L. (2023). Integrated STEM approaches and associated outcomes of K-12 student learning: A systematic review. *Education Sciences, 13*(3), 297. https://doi.org/10.3390/educsci13030297

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*(3), 681-718. https://doi.org/10.3102/0034654315627366

LiCalsi, C., Reese, K., & Garcia-Piriz, D. (2019). *Study of physical science and engineering invention kit curriculum for middle school: External evaluation of the Investing in Innovation Central Virginia Advanced Manufacturing Development Grant 78*. American Institutes for Research. https://files.eric.ed.gov/fulltext/ED611398.pdf

Lin, Q., Yin, Y., Tang, X., Hadad, R., & Zhai, X. (2020). Assessing learning in technology-rich maker activities: A systematic review of empirical research. *Computers & Education, 157*, 103944. https://doi.org/10.1016/j.compedu.2020.103944

MBADIKA. (n.d.). *M°Lab Field Kit 2.0*. Retrieved September 10, 2026, from https://www.mbadika.org/mlab-field-kit-2

Metwaly, M., Breidi, F., Garcia, J. M., & Goodridge, W. H. (2025, June). *Innovative educational kit for manufacturing: A study of student engagement and learning outcomes* [Conference paper]. 2025 ASEE Annual Conference & Exposition, Montreal, Quebec, Canada. https://doi.org/10.18260/1-2--56802

National Academy of Engineering, & National Research Council. (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research*. The National Academies Press. https://doi.org/10.17226/18612

National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits*. The National Academies Press. https://doi.org/10.17226/12190

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press. https://doi.org/10.17226/13165

Okundaye, O., Natarajarathinam, M., Qiu, S., Kuttolamadom, M. A., Chu, S., & Quek, F. (2022). Making STEM real: The design of a making-production model for hands-on STEM learning. *European Journal of Engineering Education, 47*(6), 1122-1143. https://doi.org/10.1080/03043797.2022.2121685

Reining, N., & Kauffeld, S. (2022). Empirical findings on learning success and competence development at learning factories: A scoping review. *Education Sciences, 12*(11), 769. https://doi.org/10.3390/educsci12110769

Roll, M., & Ifenthaler, D. (2021). Learning factories 4.0 in technical vocational schools: Can they foster competence development? *Empirical Research in Vocational Education and Training, 13*, Article 20. https://doi.org/10.1186/s40461-021-00124-0

Sabogal Guerrero, L. M. (2023). *Modular kit to support STEM-based electronic learning* [Undergraduate thesis, Universidad de los Andes]. Séneca: Repositorio Uniandes. http://hdl.handle.net/1992/69670

Thibaut, L., Ceuppens, S., De Loof, H., De Meester, J., Goovaerts, L., Struyf, A., Boeve-de Pauw, J., Dehaene, W., Deprez, J., De Cock, M., Hellinckx, L., Knipprath, H., Langie, G., Struyven, K., Van de Velde, D., Van Petegem, P., & Depaepe, F. (2018). Integrated STEM education: A systematic review of instructional practices in secondary education. *European Journal of STEM Education, 3*(1), Article 02. https://doi.org/10.20897/ejsteme/85525

Valjak, F., Kapetanović, A., Taradi, I., & Bojčetić, N. (2023). Work in progress: Development of educational kit for teaching additive manufacturing. *Proceedings of the Design Society, 3*, 3909-3918. https://doi.org/10.1017/pds.2023.392

Wan, Z. H., Jiang, Y., & Zhan, Y. (2021). STEM education in early childhood: A review of empirical studies. *Early Education and Development, 32*(7), 940-962. https://doi.org/10.1080/10409289.2020.1814986

What Works Clearinghouse. (2021). *Study of physical science and engineering invention kit curriculum for middle school: External evaluation of the Investing in Innovation Central Virginia Advanced Manufacturing Development Grant 78* [Study review]. Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Study/90185

Appendix A

Selected Evidence Map

The table below consolidates the strongest and most directly relevant sources used in this review. It distinguishes general learning mechanisms from direct kit evaluations, design/development studies, and practice-based examples.

Table A1

Selected Evidence and Implications for Kit Design

Source	Evidence type	Key contribution	Implication for kits
Freeman et al. (2014)	Meta-analysis; 225 studies	Active learning improved undergraduate STEM performance relative to traditional lecture.	Supports active participation broadly; not kit-specific evidence.
Lazonder & Harmsen (2016)	Meta-analysis; 72 studies	Guidance improved inquiry activity, successful performance, and learning outcomes.	Include prompts, scaffolds, feedback, and facilitation rather than relying on free exploration.
Carbonneau et al. (2013)	Meta-analysis; 55 studies; N = 7,237	Concrete manipulatives produced small-to-moderate benefits over abstract-symbol-only mathematics instruction.	Tangible representation can help, but effects vary and the evidence is adjacent rather than kit-specific.
Fyfe et al. (2014)	Systematic review	Concrete-to-abstract fading was proposed as a way to combine benefits of concrete and abstract representations.	Use the object as a bridge to concepts and generalized representations.
Davies & Seitamaa-Hakkarainen (2025)	Systematic review; 62 studies	Maker education showed varied learning and identity outcomes; science and mathematics learning was not automatic.	Design maker tasks to foreground the intended disciplinary concepts.
Elmoazen et al. (2026)	Meta-systematic review; 17 reviews	Makerspace literature reported benefits alongside curriculum, training, and equity challenges.	Treat implementation conditions as part of the intervention.
Lin et al. (2020)	Systematic review; 60 studies	Maker assessment used diverse methods; only 15% reported psychometric evidence.	Use outcome-specific measures and document their quality.
LiCalsi et al. (2019); What Works Clearinghouse (2021)	Quasi-experimental evaluation; 419 students	The invention-kit curriculum showed no statistically significant effects; implementation fidelity was uneven.	Tools alone do not guarantee impact; implementation and alignment matter.

Source	Evidence type	Key contribution	Implication for kits
Metwaly et al. (2025)	Conference study; engineering technology	A fluid-power kit connected theory with manufacturing applications and authors reported gains in understanding and engineering identity.	Promising example of modular technical practice; interpret results cautiously because evidence is course-based and nonrandomized.
Valjak et al. (2023)	Design/development study	A 17-model additive-manufacturing kit teaches foundational material-extrusion concepts before advanced AM/DfAM topics.	Physical models can provide introductory access without requiring a full equipment-based experience.
Okundaye et al. (2022)	Small-sample high-school manufacturing intervention	Micro-manufacturing integrated fabrication, production requirements, and real products; authors reported descriptive self-efficacy gains.	Authentic production constraints can connect making with manufacturing practice, but evidence is not yet generalizable.
Abele et al. (2017)	State-of-the-art review	Learning factories support practice-oriented manufacturing education, training, and research.	Provides a high-fidelity comparison for thinking about what a portable kit can and cannot approximate.
Reining & Kauffeld (2022)	Scoping review; 24 publications/22 studies	Learning factories showed competence-development promise, but evaluation rigor varied.	Specify competencies and measure them directly; avoid broad impact claims.
Roll & Ifenthaler (2021)	Repeated competence testing; N = 63	Learning Factory 4.0 participation supported subject-related technical competence but not broader digital competence.	Match claims to the specific competency targeted by the activity.
MBADIKA (n.d.); Jones et al. (2026)	Practice example and federal program report	M°Lab uses materials-science investigations in a uniform-redesign scenario; the federal report documents 21 workshops plus a session and 700+ participants.	Strong design and reach example; reach should be separated from demonstrated learning impact.
Granite Hearth (2026)	Internal instructional kit	Four Stories externalizes interviewing, coding, theme development, and memo writing.	Illustrates how a kit can make a disciplinary practice tangible even without engineering hardware.

Note. Evidence types and implications reflect the characterization used in this narrative review. Practice examples describe design or reach and are not presented as causal evidence of learning effectiveness.