

WHITE PAPER

RESPONSIVE
EVALUATION
DESIGN (RED)
A FRAMEWORK FOR
EVALUATING NSF
S-STEM PROJECTS



S-STEM REC



Acknowledgments

This white paper was developed through the Scholarships in Science, Technology, Engineering, and Mathematics (S-STEM) Resource and Evaluation Center (REC), which works to strengthen national infrastructure, build community, coordinate efforts, and synthesize findings across National Science Foundation (NSF)-funded S-STEM projects at colleges and universities across the United States.

The S-STEM REC is supported through a cooperative agreement with the American Association for the Advancement of Science (AAAS) and the National Science Foundation (NSF) under Award No. DUE-2224093. Any opinions, findings, interpretations, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of AAAS, its Board of Directors, its Council, its membership, or the National Science Foundation.

This white paper is intended to contribute to ongoing conversations about evaluation designs and methodologies in the context of S-STEM projects and to support reflection, learning, and practice. It is not a substitute for official NSF S-STEM program guidelines or the general generation of knowledge, research, or evaluation requirements. Readers are encouraged to consult the current NSF S-STEM solicitation and related program documentation for specific guidance.

Please note that while AAAS S-STEM REC promotes engagement with [MN Associates, Inc.](#) as a Resource and Evaluation Center partner, such engagement does not constitute endorsement by the National Science Foundation.

Main Authors

- Kavita Mittapalli, Ph.D.
- Nina de las Alas

MN Associates, Inc.
Evaluation Technical Assistance Providers, AAAS S-STEM REC
Website: www.mnassociatesinc.com
Primary Author Contact: kavita@mnassociatesinc.com

Credit: MN Associates, Inc. (MNA). (2026). *Responsive Evaluation Design: A Framework for Evaluating NSF S-STEM Projects. A White Paper.*
AAAS S-STEM Resource and Evaluation Center.

Table of Contents

Overview	3
Who This White Paper Is For	3
Introduction: The Case for Responsive Evaluation in S-STEM Contexts	4
Why STEM Education Could Use Responsive Evaluation	4
Responsive Evaluation Design: Principles and Framework	7
Defining Responsive Evaluation	7
Core Principles of Responsive Evaluation Design (RED)	7
RED’s Alignment with NSF S-STEM Objectives	8
Listening First: Discovering Non-Academic Barriers	9
Background Literature Review: Evaluation in STEM Education	11
Historical Context	11
Philosophical Paradigms	12
Critical Realist Evaluation of Peer Mentoring Across STEM Disciplines	12
Evaluation Frameworks and Approaches Through a Responsive Lens	13
Logic Models	13
NSF S-STEM Project Logic Model Sample	14
Adapting a Logic Model Mid-Grant: Transfer Shock Across STEM Fields	16
CDC Framework for Program Evaluation	17
Kirkpatrick Model	17
Kirkpatrick Enriched by Responsive Methods: Faculty Mentor Development	18
Utilization-Focused Evaluation (UFE)	18
RED in Action: Mock Examples from S-STEM Projects	21
Common Challenges and How RED Addresses Them	23
Ethical Considerations in Responsive Evaluation	24
Emerging Trends in Responsive Evaluation Relevant for STEM Education	25
Conclusion	27
References	28



Overview

STEM education, spanning the natural sciences, mathematics, technology, and engineering, operates within complex, dynamic, context-dependent systems where evaluation plays a vital role in understanding project effectiveness, ensuring implementation fidelity, and guiding continuous improvement. This white paper advances Responsive Evaluation Design (RED) as a promising framework for evaluating NSF S-STEM projects across all STEM disciplines.

Rooted in Robert Stake's (1975, 2004) foundational principles, RED centers stakeholder concerns, privileges contextual knowledge, and enables iterative, real-time inquiry, making it well-suited to the evolving, access-focused goals of S-STEM initiatives. Unlike static evaluation models, RED adapts as projects unfold, centering the voices of low-income scholars, faculty mentors, and institutional partners throughout the evaluation cycle.

This white paper distinguishes evaluation from (learning) assessment, presents a comparative review of other leading evaluation frameworks, including logic models, CDC Framework (2025), the Kirkpatrick Model (2006), and Utilization-Focused Evaluation (2008), and situates each within the responsive evaluation design framework and tradition. The paper employs mixed methods and realist evaluation to illustrate how evaluators can investigate not only what works but also why and under what conditions. Illustrative scenarios ground the discussion in practice. Key challenges, limitations, ethical considerations, and forward-looking trends, including AI-driven analytics and participatory methods, are addressed too.

Who This White Paper Is For

This white paper is written for anyone who designs, implements, funds, or evaluates NSF S-STEM projects.¹

This paper does not assume prior expertise or experience in social or evaluation theories or frameworks. It is accessible to PIs, co-PIs, and others in STEM departments who are reviewing or designing an evaluation plan. It is also useful to evaluators seeking an access-centered framework and to project officers seeking to better understand S-STEM evaluation design. The emphasis throughout is on honest, contextually grounded, stakeholder-centered, and genuinely useful evaluation, not evaluation as compliance, but evaluation as learning.

PRIMARY AUDIENCE	RELEVANCE TO THIS PAPER
S-STEM Principal Investigators & Co-PIs	Understand how RED can improve S-STEM projects in real-time, not just satisfy reporting requirements.
S-STEM External & Internal Evaluators	A coherent, access- and opportunity-aligned framework applicable across STEM disciplines.
NSF S-STEM Project Officers & Reviewers	As a potential framework to help understand what rigorous, responsive evaluation design looks like in S-STEM contexts.

¹The RED framework applies to any evaluation work across disciplines. However, we are contextualizing it in reference to the S-STEM grants for the white paper.

Introduction

THE CASE FOR RESPONSIVE EVALUATION IN S-STEM CONTEXTS

STEM fields lie at the core of the nation's scientific and economic enterprise. The Bureau of Labor Statistics projects STEM occupations will grow at about three times the rate of non-STEM occupations (8.1% vs 2.7%) over the next decade (BLS, 2024), spanning healthcare, environmental science, data analytics, AI, cybersecurity, and infrastructure. NSF's Scholarships in Science, Technology, Engineering, and Mathematics (S-STEM) program directly addresses these workforce imperatives by providing financial and academic support to academically talented, low-income students pursuing STEM degrees and careers.

S-STEM grants fund not only scholarships but also evidence-based curricular and co-curricular activities and supports, such as academic advising, mentoring, cohort-building, summer bridge, undergraduate interdisciplinary research, and industry and graduate school pathways, that enhance persistence and promote degree completion.² These activities, supports, and on-campus resources take different forms across projects, but share a common purpose: supporting S-STEM scholars who face financial and systemic barriers.

Evaluating these multifaceted and often interdisciplinary projects requires an approach and a framework that can accommodate context and complexity, respond to emergent findings, and center the voices of those most affected. Responsive Evaluation Design (RED) is one such promising approach.

WHY STEM EDUCATION COULD USE RESPONSIVE EVALUATION

STEM education spans a wide range of disciplinary cultures, pedagogical traditions, and institutional contexts. A biology project emphasizing field-based research apprenticeships presents very different evaluation challenges than a mathematics project focused on gateway course success, or a computer science project built around industry partnerships.

What these projects share is complexity: multiple interacting components, diverse student populations, and outcomes that unfold over years.

S-STEM projects add further context and complexity. They serve students whose educational journeys are shaped by financial insecurity, first-generation status, location, transfer pathways, family obligations and needs, and the lived experience of navigating institutions not historically designed for them. Therefore, traditional pre-ordinate evaluation models and approaches that measure fixed outcomes against predetermined benchmarks can tell us whether retention rates improved, but not why. Responsive evaluation generates a deeper contextual understanding, not just measurement, and places that understanding in the hands of those who can act on it.

We know that STEM education evaluation has historically relied on approaches grounded in accountability, outcomes measurement, and program fidelity. Experimental and quasi-experimental designs, logic-model-driven evaluations, utilization-focused evaluation, developmental evaluation, and CIPP (Context, Input, Process, Product) models are all commonly used across NSF-funded and institutional STEM initiatives. Each offers important strengths depending on the purpose of the evaluation, the maturity of the intervention, stakeholder needs, and the nature of the questions being asked.

For example, experimental and quasi-experimental methodologies are valuable when the primary goal is to establish causal relationships or demonstrate effectiveness through measurable outcomes. These are particularly useful for examining questions related to retention, persistence, GPA, graduation rates, or transfer outcomes. Similarly, logic-model-based evaluations provide structure and alignment among activities, outputs, and intended outcomes, supporting accountability and reporting requirements common in federally funded STEM projects.

However, many STEM education initiatives, particularly S-STEM projects, operate within highly dynamic institutional and socio-economic environments where outcomes are shaped by various interacting contextual factors that are difficult to isolate or standardize. Students' experiences are influenced not only by academic interventions but also by financial pressures, caregiving responsibilities, transfer pathways, mentoring relationships, institutional climate, identity development, accessible supports and resources, and ever-evolving career aspirations. In these contexts, approaches centered primarily on predetermined metrics may provide evidence of whether outcomes changed without adequately explaining how, for whom, or under what conditions those changes occurred.

Responsive Evaluation Design (RED) offers a productive and promising framework for addressing these challenges by prioritizing stakeholder perspectives, contextual understanding, and emergent learning alongside traditional outcome measurement. Rather than treating context as noise to control for, RED treats context as central to understanding program effectiveness and impact. This orientation is particularly relevant in STEM education settings where interventions often evolve, implementation varies across institutional contexts, and student needs shift in response to personal and structural realities.

RED also aligns well with the increasingly collaborative and adaptive nature of STEM education reform. Many contemporary STEM initiatives involve partnerships among faculty, student affairs professionals, industry collaborators,

community organizations, and students themselves. Responsive approaches allow evaluators to surface differing stakeholder priorities, identify unintended outcomes, and support ongoing program refinement during implementation rather than only after outcomes have been measured.

At the same time, RED involves important trade-offs. Responsive approaches can be time-intensive, require substantial stakeholder engagement, and generate findings that may be less easily standardized across sites or compared across projects. Because RED emphasizes contextual interpretation, it may not always satisfy funders or policymakers seeking highly generalizable causal claims or simplified performance metrics. In some cases, more structured or experimental approaches may be more appropriate, particularly when interventions are stable, narrowly defined, and primarily focused on testing effectiveness against specific measurable outcomes.

For this reason, this paper does not position RED as the single "best" evaluation framework for STEM education. Rather, it argues that RED represents a valuable and often underutilized approach for evaluating complex STEM education initiatives where understanding lived experiences, implementation context, adaptation, and stakeholder meaning-making are essential to interpreting outcomes.

The RED framework is especially useful when evaluation questions extend beyond "Did it work?" to include "How did it work, for whom, under what conditions, and why?"

² EBP resources: https://sstemrec.aaas.org/resources/?jsf=epro-posts&_sm=search!EBPs



Responsive Evaluation Design

PRINCIPLES AND FRAMEWORK

DEFINING RESPONSIVE EVALUATION

Responsive evaluation, as conceptualized by Robert Stake (1975), focuses on project activities rather than project intents, responds to audience requirements for information, and considers different value perspectives when reporting success and failure. Unlike evaluation models that begin with a fixed set of questions derived from project objectives, responsive evaluation begins with sustained observation and stakeholder dialogue, allowing the most important concerns to emerge organically from those closest to the project.

Stake distinguished responsive evaluation from “pre-ordinate evaluation,” which is driven by pre-specified objectives and standardized instruments. While pre-ordinate approaches are appropriate when projects are well-defined and stable, responsive evaluation is suited to projects that are complex, innovative, or serving populations whose needs are not fully understood at the outset — a precise description of most S-STEM projects.



CORE PRINCIPLES OF RESPONSIVE EVALUATION DESIGN (RED)

- 1 **Stakeholder-Centricity** / Evaluation questions emerge from the concerns and needs of scholars, faculty mentors, advisors, staff, and institutional leaders.
- 2 **Contextual Sensitivity** / The evaluation is grounded in the specific disciplinary, institutional, cultural, and social contexts in which the S-STEM project operates.
- 3 **Iterative Inquiry** / Data collection and analysis are ongoing and recursive; early findings shape subsequent inquiry.
- 4 **Emergent Design** / Methods, instruments, and foci are refined in response to what is learned as implementation unfolds.
- 5 **Holistic Portrayal** / Responsive evaluation seeks to portray the full complexity of project functioning, including unintended consequences and contested values.
- 6 **Multiple Perspectives** / Responsive evaluation deliberately represents diverse stakeholder viewpoints across STEM disciplines, backgrounds, and career stages.
- 7 **Opportunity and Access Orientation** / RED explicitly attends to access, belonging, and systemic opportunity, especially for underrepresented populations in STEM.

RED'S ALIGNMENT WITH NSF S-STEM OBJECTIVES

S-STEM projects share four core objectives that align naturally with RED. Here are a few.

S-STEM OBJECTIVE 1

DESCRIPTION	HOW RED ADDRESSES IT
Increase the number of low-income, academically talented students earning STEM degrees and entering the STEM workforce.	RED centers scholars' lived experiences and surfaces the mechanisms that support or impede degree completion.

S-STEM OBJECTIVE 2

DESCRIPTION	HOW RED ADDRESSES IT
Study the effectiveness of curricular and co-curricular activities.	Iterative inquiry and emergent design capture what works across disciplinary and institutional contexts.

S-STEM OBJECTIVE 3

DESCRIPTION	HOW RED ADDRESSES IT
Generate transferable knowledge about effective practices.	Contextual, mechanism-aware findings support cross-institutional learning beyond the funded site.

S-STEM OBJECTIVE 4

DESCRIPTION	HOW RED ADDRESSES IT
Broaden participation in STEM by supporting underrepresented groups.	Access and opportunity orientation ensures evaluation surfaces and addresses systemic disparities.

ILLUSTRATIVE SCENARIO



Listening First
DISCOVERING NON-ACADEMIC BARRIERS

An S-STEM project serving scholars in biology, chemistry, and mathematics begins with open-ended listening sessions rather than a fixed pre/post survey approach. The evaluator asks, **"What is getting in the way of your success?"** **"What do you wish the project knew about your life?"**

Across disciplines, a consistent theme emerges: the most pressing barriers are not primarily academic. Transportation costs, food insecurity, balancing coursework with employment and family caregiving, and a persistent sense of not belonging surface across all cohorts. These concerns were not part of the original evaluation framework; **they emerged from listening.**

The evaluation adapts its focus to document these systemic challenges across cohorts and grant years. Findings directly inform project discussions about emergency financial assistance, flexible scheduling, and faculty development in culturally responsive mentoring and advising. **This listening-first approach exemplifies RED's core orientation: letting what matters to stakeholders drive what the evaluation examines.**





Background Literature Review

EVALUATION IN STEM EDUCATION

HISTORICAL CONTEXT

Historically, STEM education evaluation emphasized technical outcomes measured through standardized testing and quantitative metrics — such as grades, test scores, and graduation rates — reflecting a positivist tradition that prioritized measurement over understanding (Felder & Brent, 2003). This provided accountability but frequently overlooked the complexities of learning and the diverse needs of STEM students, particularly those from underrepresented groups (NAE, 2005; ABET, 2023).

In the late 20th century, outcomes-based education (OBE) introduced measurable competencies aligned with workforce needs. ABET accreditation requires engineering programs to demonstrate how curricula prepare students for professional practice, while professional societies articulate discipline-specific learning outcomes. In the 21st century, digital technologies enable real-time feedback and richer data collection. Growing recognition of persistent inequities in STEM participation prompted a turn toward more participatory, access-centered evaluation practice. Responsive evaluation has emerged as a particularly valuable framework for this moment.

Historically, STEM education evaluation was shaped by accountability-focused traditions emphasizing standardized testing, quantitative metrics, and measurable outcomes such as grades, persistence, graduation, and workforce readiness (Felder & Brent, 2003). These approaches aligned with national concerns about scientific competitiveness and workforce development, but they often treated learning environments as uniform and paid limited attention to institutional context, student identity, or the lived experiences shaping STEM participation.

By the late 20th and early 21st centuries, several shifts began reshaping STEM evaluation practice. Outcomes-based education (OBE), accreditation movements such as ABET, and advances in educational technology expanded the use of competency-based assessment, learning analytics, and longitudinal tracking. At the same time, growing recognition of persistent inequities in STEM participation prompted evaluators to examine how financial insecurity, first-generation status, transfer pathways, family responsibilities, institutional climate, and access to mentoring influence student success, particularly in programs serving historically underrepresented populations.

These developments highlighted the limitations of relying solely on predetermined quantitative outcomes to understand program effectiveness. While traditional approaches can demonstrate whether retention or graduation rates improved, they often provide limited insight into how, for whom, or under what conditions change occurred. As STEM initiatives became more interdisciplinary, student-centered, and adaptive, evaluators increasingly turned toward approaches that foreground context, stakeholder perspectives, and implementation dynamics.

Within this landscape, Responsive Evaluation Design (RED) emerged as a particularly valuable framework for complex STEM education initiatives such as NSF S-STEM projects. RED allows evaluators to examine not only outcomes, but also the institutional, social, and personal factors shaping student experiences and program implementation. At the same time, RED is not positioned as a replacement for quantitative or outcomes-based approaches. Rather, it offers a complementary framework that is especially useful when evaluation questions extend beyond “Did it work?” to include “How did it work, for whom, and under what conditions or circumstances?”

PHILOSOPHICAL PARADIGMS

RED draws from multiple philosophical traditions, each of which contributes to its theoretical depth and methodological flexibility:

PHILOSOPHICAL TRADITION/ WAYS OF KNOWING	CORE ORIENTATION	CONTRIBUTION TO RED
POSITIVIST	Measurable outcomes: GPA, graduation rates, standardized tests/assessments (Comte, 1853; Shadish et al., 2002).	Provides accountability data and causal rigor; RED interprets these metrics within contextual data that explains them.
POST-POSITIVIST / CONSTRUCTIVIST	Integrates qualitative insights with quantitative data; centers lived experiences and meaning-making (Lincoln & Guba, 1985; Creswell & Plano Clark, 2018).	RED's insistence that evaluation represents multiple value perspectives reflects a deeply constructivist epistemology.
CRITICAL REALIST	Context→Mechanism→Outcome (CMO) configurations; investigates not only what works, but why and under what conditions. (Pawson & Tilley, 1997; Maxwell & Mittapalli, 2010; Maxwell, 2012, 2013).	CMO-informed RED can trace differential mechanisms across biology, mathematics, and technology projects rather than averaging them away.

ILLUSTRATIVE SCENARIO



Critical Realist Evaluation

OF PEER MENTORING ACROSS STEM DISCIPLINES

An S-STEM project serving scholars in biology, chemistry, and mathematics **implements a peer mentoring initiative to address documented disparities in the persistence of women and underrepresented minority scholars**. The evaluation adopts a critical realist design, asking not only whether the program improves retention, but what mechanisms drive that effect and under what conditions.

The evaluation reveals a striking disciplinary pattern: **peer mentoring is most effective in biology, where lab-based cohort work creates natural opportunities for sustained peer interaction and informal mentoring**. In mathematics, where scholars often work more independently, the formal peer mentoring structure feels artificial to both mentors and mentees, and its effects are weaker.

The mechanism at work, the creation of informal spaces where scholars can discuss shared experiences of marginalization (Yosso et al., 2009), is the same across disciplines, but the program structure facilitates it more readily in some disciplinary cultures than others. This CMO analysis leads the project team to **redesign the mathematics mentoring component around more organic peer community opportunities**: study groups, living-learning community, problem-solving sessions, and mathematical communication workshops that fit the culture of mathematical practice.

EVALUATION FRAMEWORKS AND APPROACHES THROUGH A RESPONSIVE LENS

The following four frameworks are widely used in NSF S-STEM project evaluation and are presented here through a responsive lens, illustrating how RED can enrich and extend each. These frameworks were selected because they represent the most commonly required or recommended approaches in federal STEM education evaluation contexts, and because each offer something that RED meaningfully complements: structure, accountability, outcome measurement, and utilization focus, respectively.

Logic Model Framework maps relationships among a project's inputs, activities, outputs, and intended outcomes. Through a responsive lens, the logic model becomes a living document — updated iteratively as the evaluator learns what is actually happening and used as a stakeholder reflection tool rather than a fixed accountability checklist.

CDC Evaluation Framework provides a six-step systematic process for engaging stakeholders, focusing design, gathering evidence, and ensuring use. Updated in 2024 to include five Cross-Cutting Actions emphasizing cultural responsiveness, equity, context, systems thinking, and continuous learning, the revised framework aligns closely with RED's adaptive, stakeholder-centered approach. Through a responsive lens, RED deepens implementation by building explicit iterative feedback loops between stages and centering scholar voices throughout.

Kirkpatrick Framework evaluates programs across four levels: reaction, learning, behavior, and results. Through a responsive lens, RED strengthens Kirkpatrick by surfacing the reasons behind patterns at each level, especially when aggregate outcome data cannot explain unexpected findings.

Utilization-Focused Evaluation (UFE), developed by Michael Quinn Patton, designs evaluation explicitly around the needs of identified stakeholders. RED and UFE are deeply compatible — both center actionability, with RED extending UFE by adding an explicit equity orientation and iterative inquiry.

LOGIC MODELS

Logic models are visual representations of project components, illustrating relationships among resources, activities, outputs, and outcomes (Funnell & Rogers, 2011; W.K. Kellogg Foundation, 2004). In a responsive evaluation context, logic models function as living documents — provisional theories of change that are tested, refined, and revised as the evaluation progresses and stakeholder concerns evolve.

For S-STEM projects, a logic model maps relationships among scholarship support, mentoring, undergraduate research experiences, and outcomes such as GPA, retention, and workforce entry. In a multidisciplinary project, the logic model may need to account for the possibility that the pathway from activities to outcomes looks different for biology scholars than for mathematics or technology scholars. The responsive evaluator treats the logic model as a context-driven hypothesis and the evaluation as a systematic test of that hypothesis.

ON THE FOLLOWING PAGE

NSF S-STEM Project Logic Model structured within a Responsive Evaluation Design (RED) framework. Rows depict the four interconnected program components, while columns show the traditional logic model flow from Inputs to Outcomes. Key assumptions and external factors are clearly highlighted — aligning with RED's focus on contextual transparency.

Limitation: Despite its wide recognition and usage, a logic model's often linear and rigid structure can obscure emergent outcomes and complex scholar-lived experiences that fall outside predefined pathways.

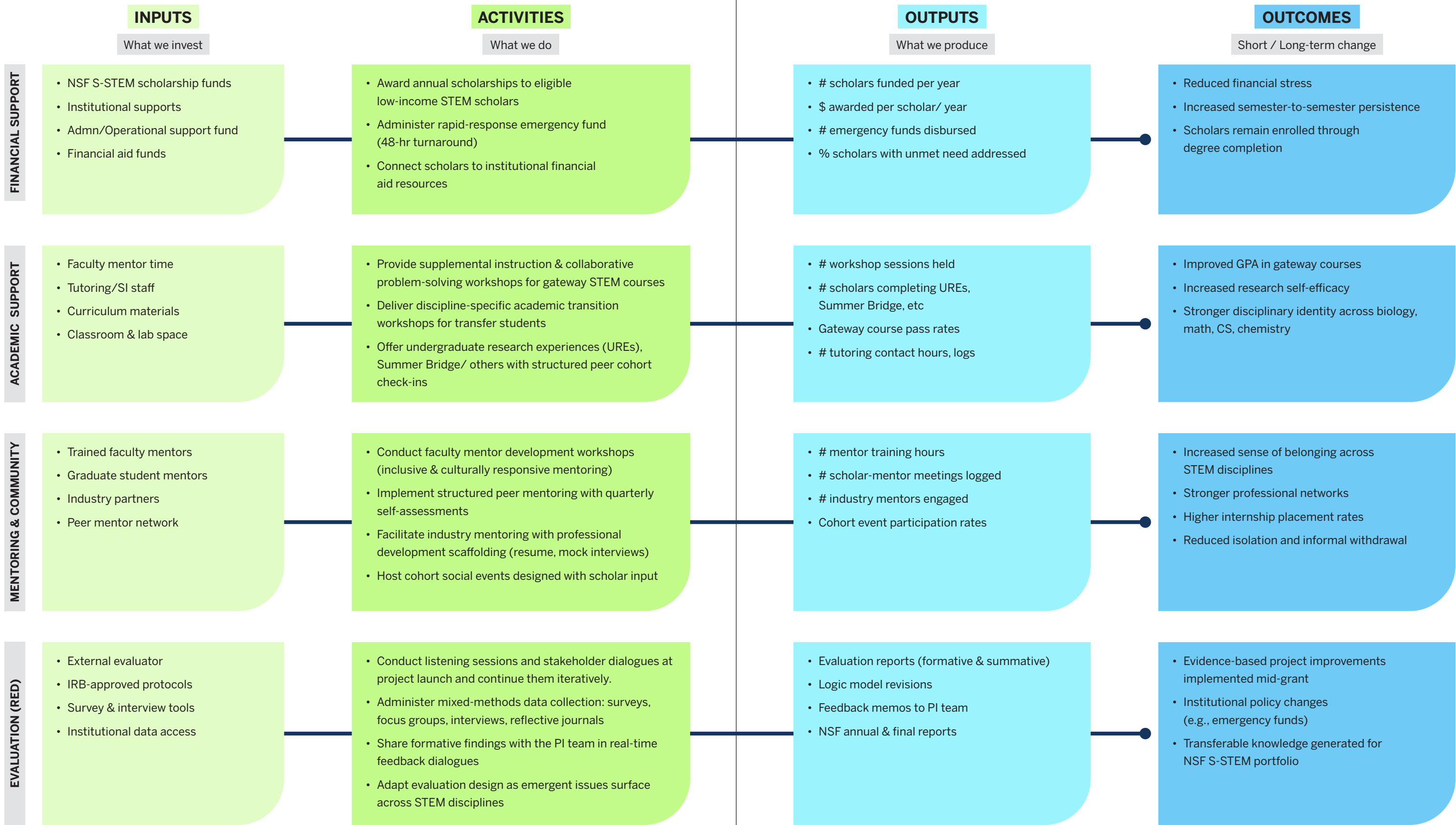
KEY ASSUMPTIONS: Scholars are academically talented and eligible per NSF criteria • Institutional leadership supports the NSF program and S-STEM project goals • Institution has support services on campus • Faculty mentors are willing to engage in professional development and training opportunities • Scholars face financial and systemic barriers that, if addressed, are surmountable

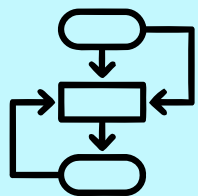
EXTERNAL FACTORS: Institutional financial aid policies • STEM labor market conditions • State & federal higher education policy • Disciplinary culture within STEM departments • Scholar family and employment obligations

LONG-TERM IMPACT: Increased number of low-income, academically talented students earning STEM degrees and entering the STEM workforce • Broadened participation in STEM • Transferable institutional and research knowledge advancing access and opportunity in STEM education



NSF S-STEM Project Logic Model Sample





Adapting a Logic Model Mid-Grant

TRANSFER SHOCK ACROSS STEM FIELDS

An S-STEM project serving community college transfer students in biology, chemistry, and mathematics uses a **logic model mapping the intended pathway from scholarship support to degree completion**. The original model assumes financial relief and cohort participation are the primary drivers of persistence.

A responsive evaluation, through focus groups and faculty advisor interviews early in Year 1, reveals an unanticipated pattern: **transfer students are struggling not primarily for financial reasons, but because of transfer shock**, the well-documented academic adjustment challenges that accompany the move from a two-year to a four-year institution (Laanan, 2001). The pattern is consistent across disciplines, though its expression differs: in chemistry, scholars report higher-than-expected laboratory demands; in mathematics, scholars report confusion about proof-writing conventions they had not previously encountered.

The project team revises the logic model mid-grant to include discipline-specific academic transition workshops. The evaluation adapts its instruments to track their relationship to first-semester performance across all three tracks. **This iterative use of the logic model as a revisable theory of change is a useful strategy** for applying RED in practice.



CDC FRAMEWORK FOR PROGRAM EVALUATION

The CDC Framework for Program Evaluation was originally introduced as a six-step approach emphasizing stakeholder engagement, program description, evaluation design, credible evidence, justified conclusions, and ensuring use and lessons learned (CDC, 1999). In 2024, the CDC updated the framework to retain these six interconnected steps while introducing five cross-cutting actions that emphasize cultural responsiveness, equity, context, systems thinking, and continuous learning throughout the evaluation process. This updated structure aligns well with Responsive Evaluation Design (RED), particularly in STEM education contexts where student experiences, institutional conditions, and implementation realities often evolve.

Within S-STEM projects, the revised CDC framework supports an adaptive and context-focused approach to evaluation. Rather than treating evaluation design as a static plan developed at project launch, evaluators revisit assumptions, stakeholder priorities and needs, and contextual conditions throughout the grant period. The inclusion of cross-cutting actions encourages evaluators to continuously assess context, engage diverse stakeholders, and integrate feedback loops that support real-time program improvement rather than relying solely on end-of-grant reporting.

A RED-informed application of the CDC framework in S-STEM settings may include iterative refinement of survey instruments, logic models, mentoring structures, or student support strategies in response to emerging evidence. This flexibility is especially important in STEM projects serving diverse student populations and disciplinary contexts, where challenges and supports may differ substantially across fields and institutions.

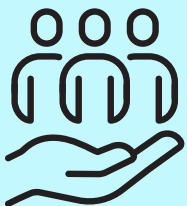
Limitation: Despite its updated flexibility, the framework remains broad by design and does not prescribe specific methods or tools, which can often challenge evaluators working in under-resourced S-STEM projects who need more concrete methodological guidance. Its comprehensive scope can also make full implementation demanding for smaller evaluation teams.

KIRKPATRICK MODEL

The Kirkpatrick Model generally evaluates training programs within S-STEM grants across four levels (Kirkpatrick & Kirkpatrick, 2006). Level 1 (Reaction) measures participants' immediate responses to a training experience: satisfaction, relevance, and engagement. Level 2 (Learning) assesses the knowledge, skills, or attitudes acquired during training. Level 3 (Behavior) examines whether participants apply what they learned in their work — the transfer of training to practice. Level 4 (Results) evaluates the broader organizational or programmatic outcomes produced by that behavior change. Applied to S-STEM faculty development workshops, mentor training, or scholar skill-building seminars, the model provides a useful ladder of evidence applicable across STEM disciplines.

Measures of impact at each level can include return on expectations (ROE) at Level 1, pre/post knowledge assessments at Level 2, observation protocols or structured self-assessments at Level 3, and return on investment (ROI) or composite return on investment (cROI) calculations at Level 4. However, in S-STEM contexts, impact is often better captured through scholar outcomes such as retention, belonging, and degree completion than through financial metrics alone. A responsive evaluator attends to how stakeholders understand and experience each level, and what unintended effects emerge.

Limitation: Its hierarchical structure prioritizes measurable results over process insights and tends toward summative rather than formative use. It does not account for the scholar outcomes directly.



Kirkpatrick Enriched by Responsive Methods

FACULTY MENTOR DEVELOPMENT

An S-STEM project conducts a **faculty mentor development workshop on inclusive mentoring**, attended by faculty from biology, mathematics, computer science, and chemistry. Kirkpatrick Level 1 and Level 2 surveys show uniformly positive responses across all disciplinary groups.

A responsive follow-up three months later reveals a striking disciplinary difference: biology faculty easily integrate inclusive mentoring into regular lab meetings, where extended one-on-one interaction is built into the work itself. Mathematics faculty describe difficulty finding natural opportunities for relational mentoring in a discipline where faculty-student contact is often brief and problem-focused.

This emergent, disciplinary finding, which would not have surfaced from Level 1–2 instruments alone, becomes the focus of a tailored follow-up session for mathematicians and other faculty in lecture-oriented disciplinary cultures. The responsive evaluator’s willingness to let Level 3 inquiry reshape the project is precisely what RED is designed to enable.

UTILIZATION-FOCUSED EVALUATION (UFE)

Utilization-Focused Evaluation (UFE), developed by Michael Q. Patton (2008, 2021), emphasizes designing evaluations for the primary intended users. UFE shares with RED a commitment to stakeholder relevance, but the two approaches differ in design orientation:

RESPONSIVE EVALUATION (STAKE, 1975)	UTILIZATION-FOCUSED EVALUATION (PATTON, 2008)
Begins with stakeholder concerns as they emerge through observation and dialogue.	Begins by identifying primary intended users and pre-negotiating how findings will be used.
Evaluation questions evolve in real-time; the evaluator is an attentive observer.	More structured upfront design, though still responsive to stakeholder needs.
Ideal for S-STEM projects with complex human dynamics, evolving goals, and access-centered missions spanning multiple disciplines.	Ideal when clear decision-makers exist and timely, targeted information is the primary evaluation goal.
In practice: interview protocols adapt after early sessions with biology scholars reveal concerns about lab safety and power dynamics not anticipated in the evaluation plan.	In practice: pre-defined outcome metrics aligned with NSF reporting requirements; dashboards so project directors can track retention across disciplinary tracks.

UFE and RED are complementary and can be productively integrated within the S-STEM context. While the UFE disciplines the evaluation toward actionable use, the RED disciplines it toward authentic stakeholder voice. Together, they can potentially produce evaluations that are both rigorous and genuinely responsive.

Limitation: Its emphasis on primary intended users can privilege project leadership perspectives over those of scholars and other less powerful stakeholders.

Although often used interchangeably, evaluation and assessment serve distinct and complementary roles in STEM education. These are adapted from the works of Scriven (1991) and Banta & Palomba (2015).

Within a responsive evaluation framework, the responsive evaluator attends not only to project-level outcomes but to the meanings that scholars, faculty, and administrators across STEM disciplines attach to those outcomes. Illustrative evaluative questions:

- To what extent does the S-STEM cohort model foster a sense of belonging and scientific identity among biology scholars, and how does this compare to the experience of mathematics scholars in the same project?
- What mechanisms support persistence among first-generation scholars in STEM gateway courses-calculus, general chemistry, introductory programming, and how do scholars describe those mechanisms?
- How are S-STEM alumni across STEM disciplines navigating the transition to a 4-year institution, graduate school, or the workforce, and what project elements do they credit or wish had been different?

Assessment vs. Evaluation

ASSESSMENT Focuses on measuring individual student learning, such as how well scholars have mastered specific knowledge, skills, and competencies.

EVALUATION Operates at the project or institutional level, seeking to understand how well a curriculum, mentoring model, or co-curricular initiative achieves its intended outcomes.

ASSESSMENT **Purpose:** feedback to scholars and instructors, supporting learning and identifying where instructional support is needed.

EVALUATION **Purpose:** judge a project’s design, implementation fidelity, and impact within its specific disciplinary and institutional context.

ASSESSMENT Looks different across STEM fields: lab reports in biology, problem sets in mathematics, coding projects in computer science.

EVALUATION Uses both qualitative and quantitative data from assessments, observations, interviews, and institutional metrics.



RED in Action

MOCK EXAMPLES FROM S-STEM PROJECTS

The following fictitious/mock examples illustrate how Responsive Evaluation Design (RED) could successfully operate in practice across S-STEM project components and STEM disciplines. Each is grounded in patterns documented in the S-STEM and STEM education evaluation literature. They are composite scenarios, not attributed to any specific institution, project, or track, and each follows a consistent three-part structure: the situation the evaluation encountered, what RED could do in response, and what can change in the project as a result.

These examples may superficially resemble standard scheduled data collection, but each illustrates something a pre-planned protocol would have missed. The evaluator's responsiveness, noticing an unexpected signal, pursuing it outside the original evaluation plan, and acting before end-of-cycle reporting, is precisely what distinguishes RED from conventional approaches. Taken together, the examples span different disciplines, project components, and types of insight: hidden stigma, subgroup dynamics, perception gaps, mid-cycle course corrections, and institutional change.

MOCK EXAMPLE 1 STEM FIELD: BIOLOGY COMPONENT: UNDERGRADUATE RESEARCH EXPERIENCE (URE)



Situation: An S-STEM project offered a summer URE designed to strengthen scientific identity and prepare scholars for graduate school. The evaluation plan focused on pre/post research self-efficacy measures and faculty mentor satisfaction. Midway through the first summer cohort, the evaluator began hearing something unexpected: several scholars felt isolated in individual lab assignments, unsure whether their struggles were normal, and reluctant to ask faculty mentors for help for fear of appearing unprepared.

RED in Action: Rather than waiting for end-of-summer survey data, the responsive evaluator convened a mid-summer focus group with the scholars; an activity not in the original evaluation plan. The group confirmed the pattern: scholars were experiencing the URE as lonely and hierarchically intimidating. The evaluator shared preliminary findings with the PI team within a week.

What Changed: The project immediately restructured the URE to include weekly peer cohort check-ins facilitated by a graduate student mentor, creating a low-stakes space for scholars to share struggles and strategies and a social for scholars on Fridays. End-of-summer reflective journals documented a marked shift, from isolated and anxiety-producing to collaborative and identity-affirming. The project retained this structural change in subsequent cohorts.

What a standard evaluation would have potentially missed: Pre/post self-efficacy surveys would have recorded modest gains at best, with no explanation for why some scholars disengaged mid-summer. The social isolation dynamic and its solution would have remained invisible until it was too late to act.

MOCK EXAMPLE 2 STEM FIELD: COMPUTER SCIENCE COMPONENT: INDUSTRY MENTORING AND INTERNSHIP PATHWAY



Situation: An S-STEM computer science project established partnerships with regional technology employers to provide industry mentors and summer internship pathways. After Year 1, internship placement rates were lower than anticipated. The project team attributed this to scholars' academic schedules conflicting with employer timelines.

RED in Action: The responsive evaluator interviewed both scholars and industry mentors to understand the gap. Industry mentors described scholars as technically strong but uncertain in professional communication. Scholars described the mentoring relationship as transactional and awkward, "like a job interview every time we meet." Neither group had raised these concerns in structured project feedback forms. The evaluator synthesized these parallel accounts: the project's mentoring model lacked a developmental scaffolding phase before scholars were expected to engage professionally with industry partners.

What Changed: The project added a structured professional development series — resume workshops, mock interviews, and networking practice — delivered by graduate student mentors before scholars were connected with industry

mentors. In the following year, internship placement rates increased, and both scholars and industry mentors reported more comfortable and productive relationships.

What a standard evaluation would have missed: Placement rate data pointed to scheduling as the culprit. Without responsive interviewing of both scholar and employer perspectives, the communication and relationship scaffolding gap would never have been identified — and the project would have continued adjusting calendars rather than addressing the real barrier.



MOCK EXAMPLE 3
STEM FIELD: PHYSICS
COMPONENT: SCHOLAR COHORT
AND BELONGING INTERVENTIONS

Situation: An S-STEM physics project built a cohort model in which scholars took two courses together each semester and participated in monthly cohort social events. End-of-year belonging survey scores were strong. Yet the evaluator noticed that two scholars had quietly withdrawn from the project mid-year and that several others had stopped participating in cohort social events.

RED in Action: The evaluator followed up with scholars who withdrew and disengaged through informal conversations, making clear that participation was voluntary and confidential. A pattern emerged: first-generation scholars described feeling socially out of place in cohort events implicitly structured around shared cultural capital — restaurant outings that were financially inaccessible, social references that felt exclusionary, and a group culture that had coalesced around a subset who had known each other before the project. The belonging survey had not captured this subgroup dynamic.

What Changed: The project redesigned its cohort programming based on scholar input, including the addition of on-campus events, potluck meals, study-oriented social activities, and a small discretionary fund to cover participation costs. The evaluator added a supplemental belonging item, asking specifically about cohort social events and peer relationships. Subsequent data showed improved engagement among first-generation scholars and reduced informal withdrawal.

What a standard evaluation would have missed: Aggregate belonging scores were strong. The project looked like it was succeeding. The evaluator’s attention to who was quietly leaving, and why, revealed a subgroup experience that the survey instrument was not designed to detect.



MOCK EXAMPLE 4
STEM FIELD: MULTI-DISCIPLINARY
(BIOLOGY, MATHEMATICS, TECHNOLOGY)
COMPONENT: FINANCIAL AID AND
EMERGENCY SUPPORT SERVICES

Situation: A multidisciplinary S-STEM project assumed that scholarship funding, along with existing institutional financial aid services, would be sufficient to address scholars’ financial needs. In Year 2, retention rates in the mathematics cohort dropped unexpectedly.

RED in Action: The responsive evaluator conducted a round of individual check-in interviews with mathematics scholars. Several had experienced mid-semester financial emergencies, including unexpected medical bills, car repairs needed for commuting, or loss of part-time employment. They also had been unable to navigate the institution’s financial aid office quickly enough. Some had considered withdrawing. The existing financial aid system was too slow and procedurally complex for emergency needs. This affected scholars across all disciplinary tracks, but it first reached a critical threshold in mathematics.

What Changed: The project worked with the institution to establish a rapid-response emergency fund, separate from formal financial aid, with a streamlined application process that could be completed within 48 hours. The evaluator documented the institutional advocacy process as part of the evaluation narrative, demonstrating how RED findings can drive institutional change beyond the S-STEM project itself. Mathematics cohort retention stabilized in the following semester.

What a standard evaluation would have missed: Retention data flagged a problem, but did not explain it. An annual survey would have recorded financial stress as a general theme at year’s end — after scholars had already left. The responsive check-in identified the mechanism in time to intervene, and the evaluator’s documentation turned a project-level finding into an institutional policy change.

COMMON CHALLENGES AND HOW RED ADDRESSES THEM

STEM education projects often operate within complex institutional, disciplinary, and resource environments that can shape both implementation and evaluation. The table below highlights several common challenges encountered in NSF S-STEM initiatives and illustrates how Responsive Evaluation Design (RED) can provide adaptive, context-sensitive strategies for addressing them. By emphasizing stakeholder engagement, flexibility, and continuous learning, RED helps evaluators respond to evolving project realities while maintaining rigor and usefulness.

CHALLENGE	DESCRIPTION	RED'S RESPONSE
Disciplinary Fragmentation	S-STEM projects spanning multiple STEM disciplines face challenges in interpreting project-level summary data that may mask significant disparities across fields.	RED treats disciplinary context as a primary variable. Instruments, protocols, and analysis are designed to surface and examine disciplinary differences.
Accreditation & Accountability Pressures	ABET for engineering, professional society standards for biology and chemistry, and regional accreditation all emphasize quantifiable outcomes that may not capture the full richness of student experience.	Accreditation metrics become one data source among many, interpreted within contextual and experiential data. RED documents how alignment with external standards shapes the learning environment for low-income scholars.
Resource Constraints	Comprehensive multi-method evaluation demands significant time, expertise, and personnel. Many S-STEM projects operate within tight budgets.	RED's flexibility can reduce evaluation costs by focusing available resources strategically on what matters most. Embedding evaluation within regular project activities (e.g., brief reflective writing, standing agenda items at mentor meetings, and contact log notes) generates rich data at minimal additional cost. These could be planned ahead to avoid additional burdens on the project and evaluator/teams.
Resistance to Change	Stakeholders may resist recommendations that challenge established disciplinary practices, from lecture-based mathematics courses to large introductory science labs.	Stakeholder engagement from the outset is RED's most powerful antidote to resistance. When faculty and administrators have shaped the evaluation questions, they are more likely to trust the findings.
Evaluator Positionality	Evaluators' social identities, disciplinary backgrounds, and institutional relationships shape what they observe and how they interpret data.	RED requires ongoing reflexivity and transparent documentation of positionality. Member checking — sharing preliminary findings with stakeholders for verification — is a particularly valuable practice, especially across disciplinary boundaries. For member checking to be meaningful, evaluators should create structured opportunities for stakeholders to challenge interpretations, not merely verify them.

ETHICAL CONSIDERATIONS IN RESPONSIVE EVALUATION

Ethical tensions arise from the need to balance accountability, transparency, and confidentiality in STEM education evaluations. Evaluators must adhere to established ethical standards (American Evaluation Association, 2025) while remaining genuinely responsive to stakeholders' needs across disciplines and institutional contexts.



FOUR ETHICAL CORNERSTONES FOR RESPONSIVE S-STEM EVALUATION

- 1

Informed Consent / Consent is not a one-time form; it is an ongoing relationship. As the evaluation adapts and new data-collection activities are added, as they will be in a genuinely responsive design, participants should be informed and given the option to consent to or decline each new activity.
- 2

Confidentiality / Protecting identities and personal information is central, particularly in small S-STEM cohorts, common in less-populated STEM disciplines, where even aggregate findings may inadvertently identify individuals. Implement robust safeguards: data anonymization, secure storage, controlled access, and careful reporting practices.
- 3

Access and opportunity in Stakeholder Engagement / Ensure scholars from every STEM field in the project have opportunities to contribute, not only those in larger or more visible tracks. Attend to intersecting dimensions of identity: the experience of a first-generation Latina biology scholar may differ substantially from that of a first-generation white male computer science scholar.
- 4

Evaluator Reflexivity / The responsive evaluator is not a neutral instrument. Ongoing reflexivity — examining how one's own perspective may shape the evaluation — and transparency about that process in reports and communications with stakeholders are ethical imperatives, not optional add-ons.

Emerging Trends

IN RESPONSIVE EVALUATION RELEVANT FOR STEM EDUCATION

Responsive evaluation does not exist in isolation; it draws from and connects to a broader ecosystem of evaluation approaches and emerging methodological trends. The following trends are particularly relevant for S-STEM projects, either because they share RED's equity orientation, complement its methods, or offer tools that strengthen its implementation. Evaluators and project teams familiar with these approaches will be better positioned to make intentional, informed decisions about evaluation design.

EMERGING TREND	RELEVANCE TO S-STEM RESPONSIVE EVALUATION
Equity-Centered & Transformative Evaluation (Mertens, 2009)	Responsive evaluation is a natural partner for transformative approaches. When concerns of low-income, first-generation, and underrepresented minority scholars are centered, findings inevitably reveal systemic issues that require systemic responses. Transformative evaluation, developed by Donna Mertens (2009), explicitly positions evaluation as a tool for social justice, centering the experiences of marginalized groups in every phase of the process.
AI & Learning Analytics (Arnold & Pistilli, 2012)	AI tools are most valuable when interpreted alongside the qualitative, contextual data generated by responsive evaluation. Predictive models identify which scholars are disengaging; RED explains why and what support is most likely to help. Learning analytics uses data from digital platforms to track student engagement, performance, and risk of attrition in real-time.
Participatory Evaluation (Cousins & Earl, 1992)	Involving scholars as co-evaluators, training them to conduct peer interviews, analyze data, and present findings, improves cultural responsiveness and provides scholars with research and analytical skills that strengthen STEM career preparation. Participatory evaluation engages program stakeholders, including students, as active partners in designing and conducting the evaluation, rather than as passive subjects
Developmental Evaluation (Patton, 2011)	Designed for programs in complex, rapidly changing environments. Like RED, it emphasizes real-time learning and adaptive management, particularly valuable for S-STEM projects piloting innovative approaches to mentoring, undergraduate research, or industry partnerships. Developmental evaluation, created by Michael Quinn Patton (2011), embeds evaluation within the program itself to support innovation and learning as the program evolves.



Conclusion

SUMMARY OF KEY POINTS

This white paper has argued for Responsive Evaluation Design (RED) as a useful and promising framework for evaluating NSF S-STEM projects. While examples throughout this paper frequently reference biology, chemistry, and mathematics — the most common S-STEM disciplines — RED's principles extend across all NSF-eligible fields, including computer science, psychology, data science, social sciences, and interdisciplinary programs. The framework's strength lies precisely in its adaptability to any disciplinary context. Grounded in Robert Stake's foundational principles and enriched by developments in mixed-methods, critical-realist, access-centered, and participatory evaluation, RED offers a comprehensive, flexible, and ethically robust approach well suited to the disciplinary complexity, population diversity, and access ambitions of S-STEM projects.

RED's core strengths — stakeholder-centricity, contextual sensitivity, iterative inquiry, emergent design, and access orientation — align directly with the goals and values of S-STEM projects across STEM fields. By centering the concerns of low-income STEM scholars, attending to specific institutional and disciplinary contexts, and generating immediately actionable knowledge, RED produces evaluations that are not just methodologically sound but genuinely useful.

RED is not without challenges. It requires evaluators with specialized training in qualitative and mixed-methods, as well as comfort navigating and appreciating various theories, and establishing and maintaining stakeholder relationships across power differentials. These skills may limit evaluator availability or increase costs. The iterative, emergent nature of RED also places ongoing demands on project staff and scholars, who must be willing to engage substantively with the evaluation process rather than simply participating in periodic surveys and focus groups/discussions. Teams should weigh these demands against the depth of insight RED provides when selecting an adequate evaluation framework.

CALL TO ACTION

As STEM education evolves to serve an increasingly diverse student population and a complex workforce landscape, so too must the methods used to evaluate programs that support that evolution. We call on evaluators, principal investigators across all STEM disciplines, and NSF program officers to embrace responsive evaluation as the guiding framework for S-STEM evaluation. Specifically, we recommend:

- Begin evaluation with sustained listening across all STEM disciplinary cohorts, open-ended dialogue with project team leads, faculty, and staff, before finalizing evaluation questions and instruments.
- Design evaluation instruments and protocols that are sensitive to disciplinary culture and context, not just program-level outcomes.
- Treat the evaluation plan as a living document, revisited and revised as the project evolves and new concerns emerge across disciplines and scholar cohorts.
- Prioritize scholar voice in all phases of evaluation, from question formulation to data interpretation to findings dissemination, with attention to disciplinary representation.
- Integrate responsive evaluation findings into real-time project decision-making through regular feedback dialogues with the PI team across STEM departments.
- Attend explicitly to access in both the evaluation process and the project practices it examines, including intersecting dimensions of identity that shape scholar experience across STEM disciplines.
- Combine responsive evaluation with complementary approaches and methodologies, mixed methods, critical-realist inquiry, participatory methods, and learning analytics, to produce evaluations that are both rich in contextual understanding and rigorous in their evidentiary standards.

References

1.

ABET (2023). Criteria for Accrediting Engineering Programs. www.abet.org

2.

American Evaluation Association. (2025). Guiding principles for evaluators. <https://www.eval.org/About/Guiding-Principles>

3.

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 267–270). ACM.

4.

Banta, T. W., Palomba, C. A., & Kinzie, J. (2015). *Assessment essentials: Planning, implementing, and improving assessment in higher education* (2nd ed.). Jossey-Bass.

5.

Bureau of Labor Statistics. (2024). Occupational Outlook Handbook: STEM occupations. U.S. Department of Labor.

6.

Centers for Disease Control and Prevention (CDC). (2024). *CDC framework for program evaluation in public health*. U.S. Department of Health and Human Services. CDC Evaluation Framework.

7.

Centers for Disease Control and Prevention (CDC). (2024). *Cross-cutting actions in the CDC evaluation framework*. U.S. Department of Health and Human Services. CDC Cross-Cutting Actions.

8.

Comte, A. (1853). *The Positive Philosophy*. Trubner & Co.

9.

Cousins, J. B., & Earl, L. M. (1992). *Participatory evaluation in education: Studies in evaluation use and organizational learning*. The Falmer Press.

10.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research* (3rd ed.). Sage.

11.

Felder, R. M., & Brent, R. (2003). Designing and teaching courses to satisfy the ABET engineering criteria. *Journal of Engineering Education*, 92(1), 7–25.

12.

Funnell, S. C., & Rogers, P. J. (2011). Purposeful program theory: Effective use of theories of change and logic models. Jossey-Bass.

13.

Greene, J. C. (2007). *Mixed methods in social inquiry*. Jossey-Bass.

14.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.). Berrett-Koehler Publishers.

15.

Laanan, F. S. (2001). Transfer student adjustment. *New Directions for Community Colleges*, 114, 5–13.

16.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

17.

Maxwell, J. A. (2012). *A realist approach for qualitative research*. Sage.

18.

Maxwell, J. A. (2013). *Qualitative research design: An interactive approach* (3rd ed.). Sage.

19.

Maxwell, J., & Mittapalli, K. (2010). *Mixed methods research*. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social & behavioral research* (2nd ed., pp. 571–597). Sage.

20.

Mertens, D. M. (2009). *Transformative research and evaluation*. Guilford Press.

21.

National Academy of Engineering. (2005). *Educating the engineer of 2020*. The National Academies Press.

22.

National Academies of Sciences, Engineering, and Medicine. (2021). *Promising practices for addressing the underrepresentation of women in science, engineering, and medicine*. The National Academies Press.

23.

NSF S-STEM Program Solicitation (NSF 25-514). National Science Foundation.

24.

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Sage.

25.

Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. Guilford Press.

26.

Patton, M. Q. (2021). *Utilization-focused evaluation* (5th ed.). Sage.

27.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage.

28.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

29.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22.

30.

Stake, R. E. (1975). *To evaluate an arts program*. In R. E. Stake (Ed.), *Evaluating the arts in education: A responsive approach* (pp. 13–31). Merrill.

31.

Stake, R. E. (2004). *Standards-based and responsive evaluation*. Sage.

32.

Strayhorn, T. L. (2012). *College students' sense of belonging: A key to educational success for all students*. Routledge.

33.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research*. Sage.

34.

W. K. Kellogg Foundation. (2004). *Logic model development guide*.

35.

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, 331(6023), 1447–1451.

36.

Weiss, C. H. (1997). Theory-based evaluation: Past, present, and future. *New Directions for Evaluation*, 76, 41–55.

37.

Yosso, T. J., Smith, W. A., Ceja, M., & Solorzano, D. G. (2009). Critical race theory, racial microaggressions, and campus racial climate for Latina/o undergraduates. *Harvard Educational Review*, 79(4), 659–691.

Note: In the preparation of this white paper, the authors used AI-assisted tools in limited and specific ways. Grammarly was used throughout the drafting process to check grammar, clarity, sentence structure and format, and style. ChatGPT was used in one narrow instance to locate and verify two citation links in the Background Literature Review section that could not otherwise be found (e.g., the CDC Evaluation Framework and Cross Cutting Actions pages). The Illustrative/Mock Examples are not AI-generated. They are drawn from MNA’s own extensive evaluation experience across numerous funded grant projects over the years, with details substantially modified and embellished so that no single institution, grant, project, track, or individual is identifiable.

28 White Paper / Responsive Evaluation Design: A Framework for Evaluating NSF S-STEM Projects

sstemrec.aaas.org 29