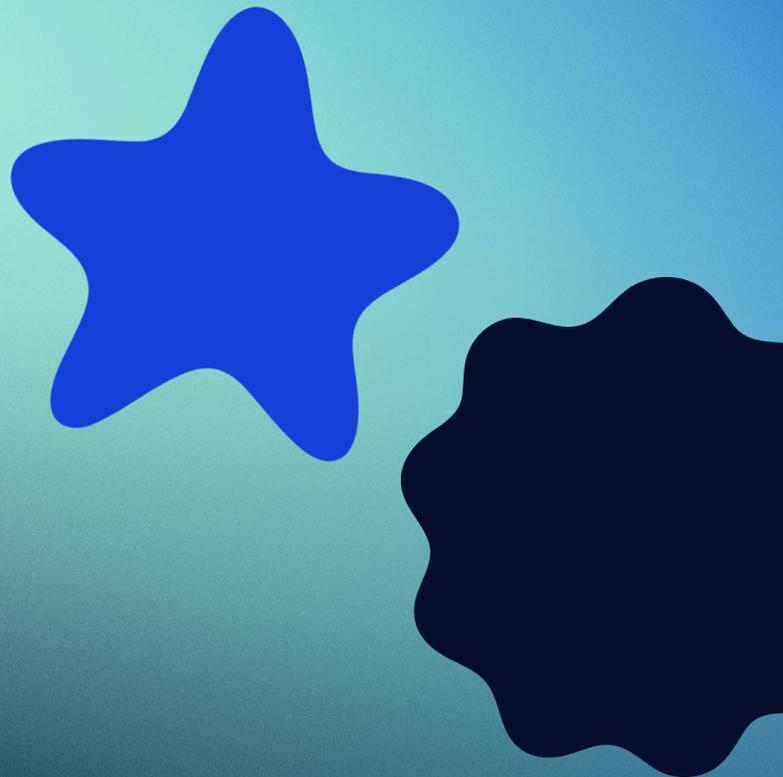


From Evidence to Implementation

A Review-Informed Programme Design
HK Unicorn Squad

Final report
2025



Report “From Evidence to Implementation: A Review-Informed Programme Design for HK Unicorn Squad” was commissioned by Proud Engineers OÜ and prepared by Stories For Impact OÜ in 2025.

Funder:

Author: Jaan Aps.

The report team gratefully acknowledges the substantive collaboration of the client's representatives and partners.

Stories For Impact partners with organisations to measure, strengthen, and communicate their impact on improving well-being and driving personal growth. We focus on healthcare, social, and education systems, as well as the social dimension of responsible business (ESG).

<https://storiesforimpact.com/en/>



Contents

Lühikokkuvõte	3
Executive summary	7
Introduction	10
Methodology	12
Literature review	12
Logic model and impact evaluation development	16
Findings	19
Factors	19
Programmes	22
Recommendations	25
Logic model and impact evaluation development	28
Logic model	28
Science-based elements checklist	30
Post-module questionnaire	35
References	38



Lühikokkuvõte

Käesoleva aruande eesmärk on koondada kokku usaldusväärsed ülevaateuringutel põhinevad tõendid selle kohta, mis toetab või takistab tüdrukute osalemist ja püsimist STEM-valdkondades. Samuti keskendutakse sellele, kuidas saadud järeldusi rakendada HK Unicorn Squadi (HKUS) programmiloogikas ja mõjuhindamise praktikas.

Töö fookus oli kolmel küsimusel:

1. millised mehhanismid ja tingimused on erinevates kontekstides seotud sellega, et tüdrukutel kujuneks seoses STEM-valdkondaga huvi, enesetõhusus ja kuuluvustunne ning nad jätkaksid STEM-aladel õppimist-töötamist;
2. millised programmi kujundamise valikud on nimetatud mehhanismide võimaldamise ja tingimuste tagamisega kooskõlas;
3. kuidas mõõta muutuseid nii, et tulemusi oleks võimalik usaldusväärsetl tõlgendada ja juhtida?

Teaduskirjanduse ülevaade viidi läbi narratiivse sünteesina; kvantitatiivset koondanalüüsi (meta-analüüsi) ei tehtud. Paralleelselt loimiti HKUSi programmi kohta tehtavatesse järeldustesse projektimeeskonna töötubades toimunud arutelude abil antropoloogilise uurimuse kvalitatiivsed tähelepanekud, et kontrollida loogikamudeli paikapidavust ning teostatavust HKUSi rakendamisel.

Teaduskirjanduse allikabaasis sisaldusid ülevaateartiklid (s.o süstemaatilised ja kaardistavad ülevaated, meta-ülevaated), mis olid avaldatud aastatel 2000–2025. Viie struktureeritud Web of Science'i päringu tulemusel saadi algkorpuseks 81 kirjet; pärast täistekstide hindamist kaasati neist 22 ning 59 jäeti põhjendatult välja. Kõik vastavad sammud dokumenteeriti, sh otsingud, sõelumine, väljaarvamine ja andmete väljavõte. Lisahindamisskaalasid ei kasutatud; kaasamiskriteeriumiks oli ülevaateartikli metoodiline läbipaistvus.

Teoreetilise alusena kasutati kahte laialt aktsepteeritud raamistikku: sotsiaalkognitiivne karjääriteooria (SCCT) rõhutab enesetõhusust, ootuskujutlusi ja nende allikaid (meisterlikkuse kogemused, eeskujud, tagasiside ja afekt), ootuste-väärtusteooria (EVT) selgitab eduootuste, ülesandeväärtuste (sisemine, utilitaarne/kogukondlik suunitlus, saavutusväärtus) ja tajatud kulude rolli püsivuses.

Kokkuvõttes selgusid viis omavahel põimunud tasandit.

1. **Psühholoogiline tasand:** tüdrukute osalemist suunavad mehhanismid on mitmetahulised; enesetõhusus, kuuluvustunne, huvi ja akadeemiline minapilt tugevnevad siis, kui tegevused pakuvad korduvalt tüdrukule endale usutavaid meisterlikkuse kogemusi, on tähenduslikud ning seotud isiklike ja kogukondlike eesmärkidega.



2. **Õpikeskkond ja pedagoogika:** programmide mõju osalejatele määrab teostuse kvaliteet; kaasav, "käed-külge" ja osalejale tagasisidevõimalusi pakkuv korraldus koos struktureeritud koostöö ja tehniliste rollide roteerimisega toetab osalust ja enesekindlust, samas kui kiirusele või võistluslikkusele orienteeritud ja juhendamata rühmatöö tekitab järjepidevuses lünki.
3. **Sotsiaalne tugi:** mentorlusvõrgustikud (sh väikese vanusevahega mentorid) ja perekondlik julgustus on seotud püsivusega; pelgalt üldised „inspiratsioonikõned“ annavad ebajärjekindlaid tulemusi juhul, kui tüdrukutel puudub tunnetus sarnasuse ja saavutatavuse kohta ning see komponent ei ole õppeprotsessiga lõimitud.
4. **Struktuurid ja õpiteed:** selged õpiteekaardid, enda proovile panemise võimalused varajase lõpliku osaluse otsuse langetamise vajaduseta ning nõustamine vähendavad väljalangust.
5. **Kultuuri tasand** (institutsionaalsed signaalid, "ruumid", naiste nähtavus) loob tausta, mis võib kas võimendada või neutraliseerida klassiruumis tehtavaid pingutusi. Kirjanduses korduvalt esile toodud takistuste hulka kuuluvad soostereotüübid ja „kaasasündinud geenius“ uskumus, androtsentriline õpikultuur, hägusad või kultuuriliselt kodeeritud õpiteed ning rollijaotused, mis jätavad tüdrukud tehnilise töö tuumast eemale.

Praktilisel tasandil joonistuvad välja **tõenduspõhised programmivalikud:**

- mitmenädalased juhendatud projektid ja leiutamispõhised ülesanded;
- struktureeritud koostöö koos rollide roteerimisega;
- nähtav õpitee ja järgnevad sammud (näitused, võistlused, praktikad);
- hübriidvormid (kooli- ja huviringipõhised harud);
- mitmetasandiline mentorlus ja kogukondlikkus;
- õpetaja juhitud suhtlus- ja turvalisuse tagamise võtted;
- vanemate kaasamine;
- keskkonna- ja poliitikatasandi toetus ("ruumid", naiste nähtavus, signaalid tüdrukutele).

Hindamise jaoks on otstarbekas seada programmide keskseteks (vahe)tulemusteks huvi, enesetõhususe, kuuluvustunde/identiteedi ning siduda neid järjepidevuse markeritega (kursuste valikud, üleminekud, püsimine). Hindamismeetodite valikus tasub eelistada segameetodeid: lühiküsitlused enne ja pärast, vaatlused ning intervjuud ja fookusgrupid koos korralduslike ja akadeemiliste näitajatega; võimaluse korral lisada ka pikemaajalisem järelmõõtmine.

Teaduskirjanduse ülevaate tulemuste praktiliseks rakendamiseks töötati välja kolm omavahel haakuvat tööriista:

1. **loogikamudel** (sisendid – tegevused – väljundid – tulemused koos SCCT/EVT mehhanismide kaardistusega ja kontekstuaalsete moderaatoritega);
2. **teaduspõhiste elementide kontrollnimed** (kas vaadeldavad teostusrutiinid on "olemas", "manualiseeritud ehk juhendisse viidud", "praktikasjuurutatud");
3. **poolaasta lõpu järelküsitlus**, mis võimaldab väikese koormusega jälgida sihitud mehhanisme igapäevases tegevuses. Koostoimes raamib loogikamudel oodatava



muutustEEKonna; kontrollnimekiri toetab ühtlast teostust ja parenduste prioriseerimist; küsitlus näitab, kas sihitud muutused mehhanismides ka tegelikult toimuvad.

Rapoertis esitatud materjalide tõlgendamisel ja kasutamisel tuleb arvestada ka mitmete piirangutega:

- toetuti Web of Science'i ülevaadetele ja nende kvaliteedile;
- teaduskirjanduse ülevaate tulemusi ei kvantifitseeritud metaanalüüsi teel (küll kasutati metaanalüüse ülevaate allikate hulgas);
- kontrollnimekirja tulemused peegeldavad HKUS töötajate ütluseid ja projekti hetkeseisu ning vajavad perioodilist ülevaatus ja võimalusel jätkuvat valideerimist kolmandate osapoolte poolt;
- küsimustik on mustandkujul ning vajab piloteerimist selguse ja vastusmustrite hindamiseks.

Hoolimata nendest piirangutest pakub raport tugeva aluse HK Unicorn Squadi teadus- ja tõenduspõhisuse sihipäraseks tõstmiseks ning läbimõeldud laiendamiseks uutesse kontekstidesse.



Executive summary

Purpose and approach

The project assembled **review-level evidence** on girls and STEM to inform HK Unicorn Squad's programme theory and evaluation plan, then integrated that evidence with qualitative insights from an independent anthropological study through joint workshops. The analytical focus was on **what reliably enables or impedes** girls' participation and persistence, **which programme features** recurrently align with those mechanisms, and **how to measure** change practically within HKUS. Quantitative pooling was not attempted; the synthesis is narrative and decision-oriented. **In total, 81 full-texts were assessed, 59 excluded with reasons, and 22 included** for synthesis.

Core findings – factors

Across reviews, outcomes are shaped by **interacting mechanisms** rather than single triggers.

Enablers cluster around: (i) **inclusive learning design** (hands-on, authentic tasks; structured collaboration; equitable turn-taking; formative feedback); (ii) **identity and self-efficacy supports** (relevance/communal framing; attainable exemplars; routine “mastery moments”); (iii) **social scaffolds** (mentoring constellations; family encouragement); (iv) **pathway clarity** (advising, navigable course maps, staged try-outs, cohorts); and (v) **supportive culture** (visibility of women; enabling institutional signals).

Barriers cluster around **stereotypes and “brilliance” beliefs, exclusionary pedagogy and climate, weak social signals, opaque or culture-coded pathways, and system-level bias.**

Format labels (e.g., club vs. course; online vs. in-person) are **neutral until enacted**; implementation quality governs impact.

Core findings – programmes

Effective designs are **theory-connected** (explicit linkage to expectancy, value, belonging, and efficacy mechanisms) and **communal in framing** (STEM for people/communities). Productive cores include **multi-week, coached, hands-on projects; structured collaboration with role rotation; advising and visible next steps; hybrid continuity** (in-school + after-school); **multi-mentor networks and cohorts; teacher-led climate work**; targeted **family engagement**; and environmental supports (spaces, visibility).

Role-model exposure is conditional: curated, attainable interactions embedded in learning show promise, whereas generic talks carry opportunity costs.



Measurement practice typically tracks **interest, self-efficacy, identity/belonging, and continuity indicators** (course-taking, retention), using **mixed methods** and, where possible, **follow-up** beyond immediate post-tests.

Products delivered

To convert evidence into practice, the project produced a **coherent trio of implementation tools** that serve distinct but complementary functions—programme theory, delivery standardisation, and routine monitoring. Together, they enable HKUS to specify hypothesised mechanisms, make delivery observable, and check whether intended changes are materialising over time.

1. **Logic model** that specifies HKUS inputs, activities, outputs, and short/medium/long-term outcomes, with explicit mapping from activities to Social-Cognitive Career Theory (SCCT) / Expectancy-Value Theory (EVT) mechanisms and moderators (enactment quality, context).
2. **Science-based elements checklist** translating evidence into observable delivery routines and status categories (whether these are “in place”, “manualised”, “adopted in practice”), supporting continuous improvement and standardisation.
3. **Post-module questionnaire** (half-year module) aligned to the logic model’s intermediate outcomes (interest, self-efficacy/mastery, safe climate, persistence, collaboration/voice¹, transfer, belonging/identity, intention to continue), designed for low burden and routine use.

Used together, the logic model **frames** expected change, the checklist **drives** consistent enactment and prioritises improvements, and the questionnaire **tests** whether mechanisms are moving in the intended direction. This alignment is the basis for a practical learning system that can travel with the programme during scale-up.

Recommendations for implementation and evaluation

The recommendations translate the evidence and toolset into **operational next steps** for sites and for the central team. They are organised to keep **design, delivery, and measurement** aligned to the same mechanisms of change and to ensure that improvement decisions are informed by both outcomes and implementation evidence.

- **Mechanism-aligned measurement:** Track identity, belonging, self-efficacy, and interest alongside behavioural continuity; pair outcomes with implementation evidence (participation balance, role rotation, feedback routines, mentoring contact; policy/space/visibility actions) to make results interpretable.
- **Equity-centred stance:** Combine learner-level supports with institutional levers (teacher development; inclusive spaces; visibility—women in STEM) to address system-level biases.

¹ Having chances to speak up, share ideas, and be heard/acknowledged.



- **Design the routine, not just the label:** Institutionalise structured collaboration, rotating technical roles, guided participation, formative feedback, and multi-week authentic projects; make pathways explicit and engage families early.
- **Use role-model components conditionally:** Curate for attainability/similarity, embed in mentoring constellations, instrument the mechanism, and compare against alternative uses of time.
- **Counter stereotypes across channels:** Integrate counter-stereotypic cues and communal purpose into materials, spaces, facilitation, and narratives.
- **Prioritise content + pedagogy pairs:** Deliver computational thinking and invention education through inclusive, feedback-rich designs adapted to disciplinary contexts.
- **Plan for interactions and context:** Expect non-linear, context-dependent effects; monitor combined impacts and site-level moderators.
- **Sustain hybrid formats and continuity:** Blend settings, sequence visible next steps (e.g., showcases, internships), minimise early lock-in via try-outs (short tasters before specialising), and track transitions/retention across stages.

Taken as a whole, these recommendations provide a **roadmap from evidence to action**: they specify where to focus, how to enact daily practices, and what to measure for credible learning. Implemented alongside the logic model, checklist, and questionnaire, they support an iterative improvement cycle within existing sites and provide safeguards for **responsible adaptation** in new contexts.



Introduction

Girls' participation and persistence in STEM remain uneven across settings despite sustained investment in programmes aimed at broadening access and interest. HK Unicorn Squad (HKUS) is a long-running, girls-only initiative that combines hands-on projects with multi-year continuity. This report documents an applied research effort to (i) consolidate dependable, review-level evidence on the determinants of girls' engagement and on programme features associated with stronger participation and continuity, and (ii) translate that evidence—together with locally generated qualitative insights—into a science-based logic model and a practical monitoring and evaluation toolkit for HKUS.

The evidence base was intentionally restricted to **review articles** (systematic reviews, scoping reviews, meta-reviews, structured overviews) published between 2000 and 2025. The objective was not to conduct a meta-analysis but to produce a **narrative synthesis** suitable for programme (re)design and measurement planning. Five structured Web of Science searches produced the initial corpus; after screening and full-text assessment, **22 review articles** were included and **59** were excluded with documented reasons (full-texts assessed: **81**). All search strings, eligibility rules, screening decisions, and extractions were recorded in a multi-sheet Excel workbook. No additional quality-scoring instruments were applied beyond clear, reproducible review methods as inclusion thresholds, and no effect sizes were computed.

Two complementary theoretical frames guide the analysis and the design of measurement: **Social-Cognitive Career Theory (SCCT)**—emphasising self-efficacy, outcome expectations, and their sources (mastery, vicarious experience, feedback/affect)—and **Expectancy-Value Theory (EVT)**—emphasising expectancies of success, task values (intrinsic, utility/communal, attainment), and perceived costs. These frameworks align with HKUS's age range and delivery model and are operationalised here as observable delivery elements (checklist items) and mechanism-aligned indicators (questionnaire constructs). To ensure local validity and feasibility, the literature synthesis was triangulated with an independent anthropological study of HKUS conducted in parallel; insights from interviews and observations informed early usability and relevance checks and prioritisation of deliverable programme routines (e.g., participation balance, feedback practices, parent touchpoints).

The report is organised to move from evidence to implementation. **Methods** describe the review approach and the translation pathway from evidence to programme artefacts. **Findings—Factors** synthesise psychological, pedagogical, social, structural, and cultural determinants of participation and continuity. **Findings—Programmes** distil recurring design principles, enactment patterns, and measurement portfolios across reviewed entries. **Recommendations** consolidate implications for implementation and monitoring, keeping design, delivery, and evaluation aligned to the same mechanisms of change. **Logic Model and Impact**



Evaluation Development presents HKUS’s logic model together with a science-based elements checklist and a brief post-module (half-year) questionnaire, including guidance for use.

In summary, this report provides (a) a consolidated, theory-anchored account of what reliably enables or impedes girls’ participation and continuity in STEM, and (b) an integrated set of tools—logic model, checklist, and questionnaire—that make those mechanisms testable in routine practice. The approach is designed to support disciplined improvement within current sites and responsible adaptation in new contexts, while acknowledging the scope conditions of a review-of-reviews (dependence on published syntheses, narrative rather than quantitative pooling) and the need for ongoing piloting of indicators during implementation.



Methodology

Literature review

Review purpose and scope

This review synthesises **review publications**—systematic reviews, scoping reviews, meta-reviews, and other structured overviews—on **girls and STEM**. The objective was to assemble dependable messages from review-level evidence about:

1. barriers and facilitators for girls' participation, persistence, and achievement in STEM;
2. program components and approaches reported as promising for girls; and
3. practical planning and measurement considerations (e.g., what outcomes are commonly tracked and how they are defined), to support logic-model development and impact evaluation for the HK Unicorn Squad programme.

No effect sizes were calculated and no statistical pooling was undertaken. The goal was a **narrative synthesis** of review-level findings that are usable in programme design and evaluation planning.

Governance, protocol, and data management

All operational steps—search logging, screening decisions, exclusion reasons, and data extraction—were maintained in an **Excel workbook** (multiple worksheets). Column definitions in the workbook were preserved. When two reviewers reached different preliminary judgements, the discrepancy was resolved in discussion, and the final decision was recorded in the relevant worksheet. No numerical meta-analytic procedures were introduced at any stage.

Eligibility criteria

To align the evidence base with the report's design-and-evaluation aims, inclusion criteria were defined in advance and applied consistently across screening stages. The focus was on high-level syntheses that could credibly inform programme mechanisms and measurement for school-age girls in STEM; where broader gender reviews were used, transferability to this population was assessed explicitly.

Publication type. Review articles only (systematic reviews, scoping reviews, meta-reviews, and other reviews with reproducible methods).

Topic. STEM education with a focus on **girls**; reviews addressing **gender in STEM** were eligible when their insights were reasonably transferable to girls.



Population and setting. School-age learners (primary to upper-secondary/college-entry). Evidence centred exclusively on adult women in higher education or the labour market was not included in the main synthesis unless the review contained clearly transferable implications for school-age girls.

Outcomes. Interest in STEM, self-confidence/self-efficacy, identity/belonging, participation and persistence, skill development/learning, and evaluation-relevant outcomes and indicators.

Time window. Reviews published **1 January 2000 to 30 May 2025** (as set in the Web of Science searches).

Exclusion criteria. Items were excluded if they:

- were not review articles;
- described review methods insufficiently for reproducibility;
- had an overly narrow contextual focus that limited generalisation;
- concerned mainly adult women without clear relevance for school-age girls; or
- addressed topics too specialised to inform programme design or evaluation for girls in STEM.

Information source

The bibliographic source was **Web of Science Core Collection**. Five structured searches (Searches 1–5) combined STEM terms with **review-type filters** and **girls/gender terms**; two of the searches added **planning/measurement keywords** (e.g., logic model, theory of change, evaluation, measurement) to capture review-level guidance on programme description and outcome tracking. The queries, date limits, and run order were documented.

Records from all searches were collated in a single Excel workbook and de-duplicated using title/author/year/DOI matching. Screening proceeded in two stages against predefined eligibility criteria: first at title/abstract level, then at full text.

At the full-text stage, 81 review articles were assessed. Of these, 59 were excluded with a documented reason category. The remaining 22 review articles met all inclusion criteria and were included in the narrative synthesis.

Data collection (extraction)

Data extraction was performed in the Excel workbook using a **single, pre-structured template** applied consistently to all **22 included reviews**. Each review was coded across the following fields:

- **Identification and workflow fields:**
Authors (year) (entered under *Background*), **File name**, **Reviewer 1**, **Reviewer 2**, and **“Especially relevant for this review?”** (a yes/no plus short justification used to flag direct relevance to the HK Unicorn Squad use case).



- **Descriptive methodology fields:**
Review type (e.g., systematic review, scoping review, meta-review), **Abstract** (verbatim summary for quick reference), **Main aims of the review**, and **Methods used in the review** (how the review searched, screened, and synthesised, as reported by the authors).
- **Scope and population fields (as reported by the review):**
Background and context (e.g., educational stage, disciplinary span, setting), and **Population and sample** (age bands, gender focus, and any subgroup notes provided in the review).
- **Coverage indicators transcribed from the review:**
Number of studies (number of primary studies included in the review's corpus, if reported) and **Years covered** (start–end years or textual description, as stated by the review).
- **Structured coding of findings – factors (four subfields):**
Neutral factors, Enablers, Barriers, and Cultural factors. These cells capture how the review characterises determinants related to girls and STEM (e.g., neutral/ambivalent findings; factors that facilitate interest, self-efficacy, identity/belonging, participation/persistence, or learning; factors that impede these outcomes; contextual or cultural considerations noted by the review). Text in these fields reflects the language of the source review; no reclassification beyond this three-part frame was performed.
- **Structured coding of findings – programmes (two subfields):**
Program design and Program measurement. These cells capture (i) programme components and delivery features (e.g., hands-on projects, role-model exposure, mentoring, teacher capacity building, family engagement, duration/intensity); (ii) any outcome constructs, indicators, or measurement approaches the review associates with those programmes. Content was taken from the review narrative/tables; no new instruments or scoring were introduced.
- **Synthesis-at-a-glance field:**
Authors' conclusions/recommendations, recording the review's own concluding messages most relevant for programme theory, logic modelling, and practical evaluation planning.

How tables were handled. When a review provided structured tables (e.g., characteristics of its included primary studies, typical intervention features, or common outcome constructs), the team transcribed the fields most pertinent to programme design and evaluation planning into the appropriate template cells (typically within *Program design*, *Program measurement*, and the coverage indicators *Number of studies* and *Years covered*). Table formats were not standardised across reviews beyond the fixed template headings, and no new metrics were derived.

What was not done. No effect sizes were calculated, no statistical synthesis was conducted, and no additional quality scoring tools were applied beyond what is reflected in the inclusion/exclusion rules. All entries in the extraction sheet reflect information **reported by the review itself** and mapped to the predefined fields listed above.



Synthesis methods

Given the variation in scope and emphasis across the **22** reviews, findings were integrated through a **structured narrative synthesis**. The synthesis was organised around three decision-relevant lenses:

1. **Barriers and facilitators** relevant to girls' engagement and persistence in STEM (e.g., identity, self-efficacy, belonging, role-model exposure, access, learning environment);
2. **Programme components and approaches** reported as promising (e.g., hands-on project work, authentic problems, longer-term engagement rather than one-off events, mentoring/peer support, role models, teacher capacity building, and community building); and
3. **Planning and measurement considerations** (e.g., how reviews describe outcomes such as interest, self-efficacy, and belonging; examples of indicators and instruments; and implications for a logic model and data-collection plan).

All **22** included reviews contributed to at least one of these lenses. The synthesis describes areas of convergence across reviews and notes scope limitations when conclusions are context-specific.

Risk-of-bias and certainty considerations

No independent appraisal scales or statistical diagnostics were introduced. Methodological safeguards were implemented through **inclusion thresholds** (reviews needed to state their methods clearly) and **exclusion categories** (e.g., non-review, unclear methods, or overly narrow focus). This preserved methodological discipline while avoiding additional instruments that were not part of the original plan.

Deviations from plan

The review remained consistent with its initial intent: review-level evidence only; no quantitative pooling; narrative synthesis for programme design and evaluation planning. Minor clarifications to exclusion categories (wording) were recorded in the screening sheet. No statistical procedures were introduced at any point.

Limitations

The review draws exclusively on **Web of Science** indexing (five structured searches). Relevant reviews indexed elsewhere may not have been captured. As an overview of reviews, the synthesis depends on the scope and reporting quality of the included reviews; gaps in those reviews cannot be filled from primary studies here.



Logic model and impact evaluation development

Design of the methodological workflow

This stage translated review-level evidence into a **testable programme logic** and a **practical monitoring plan** for HK Unicorn Squad. Workstreams were intentionally integrated: (i) evidence synthesis from the review of reviews; (ii) co-development with the Unicorn team through workshops and virtual meetings; and (iii) **triangulation with an independent anthropological qualitative study** reported in *HK Unicorn Squad uuringu lõpuraport* (Jaanits et al., 2025). Primary outputs were a logic model, a science-based elements checklist (Word + visual Excel), and a short follow-up questionnaire for routine use.

Data sources and inputs

1. **Review-level evidence.** Findings on mechanisms (identity, belonging, self-efficacy, interest), pedagogical levers (structured collaboration, feedback, role rotation), and pathway supports informed the causal propositions in the logic model.
2. **Anthropological qualitative study.** Interviews, observations (sessions and summer camp), and group discussions provided context: perceived value, delivery conditions, and risks. These insights were used to test face validity and local feasibility of candidate mechanisms and routines (Jaanits et al., 2025).
3. **Programme artefacts.** The internal slide deck and programme specifications (inputs, activities, outputs, science-based assumptions) anchored the Activities/Outputs tier and bounded feasibility for indicators (e.g., session dose, instructor:girl ratios, token system).

Process and products

1) Evidence-informed logic model (draft → validated)

A first draft documented: problem and target group; inputs (core team, instructors, manuals, equipment boxes, platform); activities (recruitment/support of instructors, six-module sessions, equipment logistics, camps, communications); outputs (sessions, participants, attendance, token distribution); and outcome/impact signals aligned with intermediate psychological mechanisms. The draft also distinguished **assumptions already in place** (early reach; identity-safe girls-only environment; manualised STEM content) from **assumptions to strengthen** (systematic self-efficacy and outcome-expectation shaping; standardised instructor pedagogy).

2) Co-development workshops and virtual meetings

Joint sessions with the Unicorn team—and, where relevant, alongside the anthropological team—were used to: (a) confirm that the model and measures made practical sense in day-to-day delivery; (b) agree on shared terminology so the same ideas were named the same way across all workstreams; (c) prioritise near-term refinements (reflection prompts, role



rotation, parent touchpoints); and (d) agree which routines to standardise versus leave locally flexible.

3) “Logic Model Science-Based Elements Checklist” (Word)

From the review corpus and the two organising theories (below), we derived **concrete, observable elements** with four status descriptors per item: *In place (now)*, *Manualised*, *Adoption*, and *Notes* (evidence/operational remarks). Items span critical timing, identity-safe environment, duration/continuity, learning experiences that yield mastery/efficacy, exposure and preparation, family/pathway supports, outputs & reach, quick outcome signals, and instructor capability/support.

4) Joint rating with the Unicorn team

Each checklist item was reviewed and rated together. The baseline shows clear strengths—e.g., girls-only delivery, early reach, multi-year continuity, session content manuals, instructor:girl ratio guardrails—and explicit development priorities—e.g., general facilitation checklist, inclusive pedagogy micro-skills, math “mini-wins” scaffolds, parent scripts, transition bridges, micro-questionnaires.

5) Visual consolidation (Excel)

The Word checklist was rendered into a **colour-coded status view** (*Logic Model Science-Based Elements – Status + Development priorities*), included in the report. The colours mirror the jointly agreed Word ratings.

6) Follow-up questionnaire (routine monitoring)

A short, age-appropriate post-module instrument was drafted for phones/tablets to capture **mechanism-aligned signals** (interest, self-efficacy/mastery, safe climate, persistence despite difficulty, collaboration/voice, transfer to home/school, belonging/identity, intention to continue). Response options use simple faces/labels to minimise burden and integrate into session close-out.

Theoretical foundations and justification

The logic model and checklist was anchored in **Social-Cognitive Career Theory (SCCT)** and **Expectancy-Value Theory (EVT)** for three reasons:

1. **Construct fit with programme aims and observed experience.** SCCT foregrounds **self-efficacy** and **outcome expectations** as proximal drivers of interest and choice; EVT explains how **expectancies of success** and **task values** (intrinsic/attainment/utility) and **perceived costs** shape persistence. The qualitative material echoes these levers—confidence to try, productive struggle without stigma, meaningful hands-on tasks, collaboration/voice, belonging—supporting SCCT/EVT as organising theories for design and measurement (Jaanits et al., 2025).



2. **Age-stage suitability.** Both frameworks suit late-childhood/early-adolescent transitions, aligning with Unicorn's **early reach and continuity** strategy and motivating longitudinal checks of belief/value change across modules and years.
3. **Operationalisability.** SCCT/EVT translate directly into **checklist items and indicators** (mastery moments; specific feedback; similarity/attainability via near-peers; visible utility/communal purpose; routines that normalise effort and manage perceived cost). These are traceable in the checklist and questionnaire, enabling iterative improvement cycles.

Roles, governance, and version control

- The **researcher** drafted the logic model, checklist, and questionnaire;
- the **Unicorn team** validated feasibility, supplied delivery constraints, and co-rated checklist items;
- the **anthropological team** supplied an **independent qualitative evidence stream** (*HK Unicorn Squad uuringu lõpuraport*) used for triangulation and prioritisation (Jaanits et al., 2025);
- artefacts progressed through labelled versions (Draft → Workshop → Post-workshop Revision → Report Inclusion).

Quality assurance and triangulation

- **Convergence:** items advanced only when supported by review evidence, feasible in delivery, and consistent with qualitative themes (Jaanits et al., 2025).
- **Feasibility:** indicators/routines were screened for age appropriateness, burden, and fit with session flow (e.g., 1–2-minute micro-surveys; observation aligned with instructor routines).
- **Transparency:** the checklist makes status explicit (*in place, manualised, adoption*) to target the right gap (design vs. standard vs. uptake).

Ethics and context

This part relied on anonymised, aggregate insights from the anthropological study (qualitative methods with minors under established good-practice safeguards) and focused our own work on programme-design artefacts and non-identifiable monitoring tools. Questionnaire items avoid sensitive personal data, minimise burden, and allow opt-out (Jaanits et al., 2025).

Method limitations

The logic model articulates **hypotheses to be tested** and does not constitute proven causal claims. The checklist ratings reflect joint professional judgements at the time of assessment and therefore warrant periodic re-review. The follow-up questionnaire is **formative** and requires piloting—both for clarity and for response patterns—before any broader roll-out.



Findings

Factors

This section synthesises factor-level findings from the reviewed literature, organised across five interacting layers: psychological mechanisms; learning environment and pedagogy; social supports; structures and pathways; and cultural/contextual conditions. The evidence consistently indicates that girls' participation and persistence in STEM are shaped by cumulative experiences rather than single triggers, and that outcomes emerge from interactions among mechanisms, instructional design, social signals, and local context.

1) Psychological mechanisms shaping participation: identity, self-efficacy, interest, and self-concept

Across the corpus, psychological mechanisms are multi-determinant and develop cumulatively rather than through a single trigger. Reviews converge that identity/belonging, self-efficacy (confidence in capability), interest/motivation, and academic self-concept interact with the learning context and social signals. **Neutral/conditional patterns** recur: single-cause explanations are inadequate; effects depend on how experiences accumulate and on local context (Acuña et al., 2022; Bong & Chen, 2024; van Tuijl & van der Molen, 2016). Relationships between self-concept and performance are not universally linear and vary with instructional design, peer climate, and measurement approach (Wang & Yu, 2023; Bong & Chen, 2024).

Within this landscape, **enablers** include practices that explicitly develop self-efficacy and identity—feedback that normalises struggle, meaningful success opportunities, and relevance linking—which are associated with stronger interest and persistence signals for girls (Reinhold et al., 2018; Beck, M. et al., 2021; Prieto-Rodríguez et al., 2020; Verdugo-Castro et al., 2022; Ha et al., 2023). The perceived **similarity and attainability** of female role models also enhance motivational impact (Gladstone & Cimpian, 2021; Prieto-Rodríguez et al., 2020).

By contrast, **barriers** include stereotypes that STEM requires innate “brilliance” and androcentric field images; both depress self-assessment, interest, and willingness to enter or remain in advanced tracks (Gladstone & Cimpian, 2021; Acuña et al., 2022; Beck, M. et al., 2021; Wang & Yu, 2023). Perceived misfit, micro-exclusions, and assignment to auxiliary roles (e.g., note-taking) further undermine identity and self-efficacy (Acuña et al., 2022; Beck, M. et al., 2021; Boucher et al., 2017).

2) Learning environment and pedagogy: design features that matter

Findings consistently prioritise **how learning is organised and facilitated**. **Enabling conditions** include hands-on, project-based, and authentic problem-solving formats paired with **structured collaboration, equitable turn-taking, and formative feedback**; these combinations are repeatedly linked to higher participation, confidence, and persistence among girls (Reinhold et



al., 2018; Beck, M. et al., 2021; Torres–Torres et al., 2024; Ha et al., 2023; Prieto–Rodriguez et al., 2020). Intentional routines that surface and reduce participation gaps—such as **rotating technical roles** and **explicit inclusion prompts**—strengthen outcomes (Beck, M. et al., 2021; Reinhold et al., 2018).

At the same time, **neutral factors** include the observation that format labels per se (formal vs. informal; club vs. course; online vs. in-person) are **not inherently beneficial or harmful**; the **quality of enactment** governs impact (Beck, M. et al., 2021; Bong & Chen, 2024). **Barriers** are observed in didactic, speed-oriented, or competitive climates and in **unstructured group work** that permits dominance patterns, which reduce girls’ opportunities to practise core tasks and to be recognised for technical contributions (Beck, M. et al., 2021; Boucher et al., 2017). Signals that a class or field is **male-typed** also lower belonging and course-taking intention (Wang & Yu, 2023; Acuña et al., 2022).

3) Social supports: role models, mentoring, family, and peers

Evidence reviews place **social modelling and sustained encouragement** among the strongest enabling conditions. **Relatable female role models**—perceived as competent and accessible—are linked with increased motivation, identity, and persistence intentions (Prieto–Rodriguez et al., 2020; Gladstone & Cimpian, 2021; Verdugo–Castro et al., 2022; Reinhold et al., 2018). **Mentoring** and ongoing adult guidance, including teacher encouragement, help girls navigate challenges and remain engaged (Beck, M. et al., 2021; Verdugo–Castro et al., 2022; Reinhold et al., 2018). **Family stimulation and support** through materials, conversations, and expectations further strengthen early confidence and later continuity (Avolio et al., 2020; Plasman et al., 2021).

Nevertheless, **neutral/conditional effects** appear for role–model exposure: **perceived similarity, pathway clarity, and attainable exemplars** moderate impact (Gladstone & Cimpian, 2021; Prieto–Rodriguez et al., 2020). Persistent **barriers** include the **absence of visible role models, adult under-estimation, and peer climates assigning auxiliary roles** rather than central technical responsibilities (Acuña et al., 2022; Beck, M. et al., 2021; Avolio et al., 2020).

4) Structures and pathways: advising, course sequences, and transitions

Structural features set the terms for choice at key decision points. **Enabling arrangements** include **transparent, navigable pathways**—clear course maps, **opportunities to try sub-fields without early lock-in**, and **advising that connects interests to concrete sequences**—which support continued participation (Plasman et al., 2021; Verdugo–Castro et al., 2022; Prieto–Rodriguez et al., 2020; Pérez–Felkner et al., 2025). **Cohort or learning-community models** can mitigate isolation by creating a critical mass of peers (Beck, M. et al., 2021).

There are also **neutral/mixed patterns** across different pathway configurations, with outcomes moderated by **local resources and advising norms** (van Tuijl & van der Molen, 2016; Bong & Chen, 2024). **Barriers** are observed where sequences are **opaque, high-stakes, or culture-coded** into “hard” vs. “soft” STEM tracks and where access to **advanced options** is limited; these conditions contribute to drop-off (Wang & Yu, 2023; Boucher et al., 2017).



5) Cultural and contextual conditions: norms, narratives, and system signals

Contextual conditions frame the meaning of programme experiences and influence their durability. **Enabling contexts** normalise **gender equity**, make **women's STEM contributions visible**, and provide **institutional support** for inclusive practices (Prieto-Rodríguez et al., 2020; Verdugo-Castro et al., 2022; Acuña et al., 2022). Cross-country syntheses indicate that effectiveness does not follow a simple developed–developing divide; **design and implementation quality** are more predictive than macro labels (Yu et al., 2024).

Neutral/conditional features include cross-cultural differences in **self-reporting** (e.g., modesty norms), which complicate direct comparisons of confidence or interest (Wang & Yu, 2023). System-level **barriers** persist in the form of **androcentric STEM cultures**, **enduring stereotypes**, and **exclusionary norms** embedded in institutions (Beck, M. et al., 2021; Boucher et al., 2017; Msambwa et al., 2024a; Msambwa et al., 2024b; Beroíza-Valenzuela & Salas-Guzmán, 2024). In some contexts, field–gender images differ substantially—for example, **computer science** may be presented as suitable or neutral—underscoring the importance of **local cultural mapping** before scale-up (Acuña et al., 2022).

Integrated factor map (blended to reduce overlap)

When the five layers are considered together, recurrent **enablers** centre on **inclusive learning design**—hands-on, collaborative, feedback-rich, authentic tasks delivered with **explicit participation routines**—alongside **identity and self-efficacy supports** such as relevance linking, **attainable role models**, and **structured opportunities for success** (Reinhold et al., 2018; Beck, M. et al., 2021; Torres-Torres et al., 2024; Ha et al., 2023; Prieto-Rodríguez et al., 2020; Gladstone & Cimpian, 2021; Verdugo-Castro et al., 2022). Complementary **social scaffolds**, including mentoring and consistent **family encouragement**, are associated with continuity (Avolio et al., 2020; Beck, M. et al., 2021; Verdugo-Castro et al., 2022). **Pathway clarity**—advising, navigable course maps, and **cohort models** that reduce isolation—supports sustained participation (Plasman et al., 2021; Prieto-Rodríguez et al., 2020; Pérez-Felkner et al., 2025; Beck, M. et al., 2021). At the cultural level, the **visibility of women**, **equity norms**, and **enabling institutional signals** function as supportive background conditions (Acuña et al., 2022; Prieto-Rodríguez et al., 2020; Verdugo-Castro et al., 2022; Yu et al., 2024).

By contrast, dominant **barriers** cluster around **stereotypes and “brilliance” beliefs**; **exclusionary pedagogy and climate**; **weak social or role-model signals**; **opaque pathways and culture-coded tracks**; and **system-level bias** embedded in institutions (Gladstone & Cimpian, 2021; Acuña et al., 2022; Beck, M. et al., 2021; Boucher et al., 2017; Wang & Yu, 2023; Msambwa et al., 2024a; Msambwa et al., 2024b; Beroíza-Valenzuela & Salas-Guzmán, 2024). Three interpretation guidelines recur across sources: outcomes arise from **interacting mechanisms and contexts**, implying **contingent and often non-linear effects** (Acuña et al., 2022; Bong & Chen, 2024; van Tuijl & van der Molen, 2016; Wang & Yu, 2023); **format labels do not determine impact**, which depends on **design quality and facilitation** (Beck, M. et al., 2021; Bong & Chen, 2024); and **cross-cultural**



variation in field images and reporting norms argues for **local cultural mapping** before scale-up (Acuña et al., 2022; Wang & Yu, 2023; Yu et al., 2024).

Conclusion

Taken together, these factor-level findings provide an integrated account of **what needs to shift** (mechanisms and conditions), **where levers are located** (pedagogy, social scaffolds, pathways, and culture), and **why effects differ across sites** (context and enactment). They supply the conceptual basis for the programme design choices and the measurement portfolio articulated in the subsequent sections of the report, while preserving the nuance that change is **contingent, cumulative, and context-sensitive**.

Programmes

This section synthesises programme-level findings from the reviewed entries, focusing on design principles, enactment patterns, and measurement portfolios, and concluding with practical implications for aligning design and evaluation. The emphasis is on how recurrent programme features connect to motivation and identity mechanisms, how those features are enacted in day-to-day practice, and how evidence is typically gathered to verify change. Highlights of **enablers** and **barriers/risks** are signposted in bold within each subsection to aid reading.

Design principles that recur across entries

A consistent theme is **theory-connected design**. Multiple entries argue that programmes work best when activity choices are explicitly connected to motivation and identity mechanisms—expectancies, values, belonging, and self-efficacy—while also noting that many interventions still lack a clear theoretical spine (Acuña et al., 2022; Bong and Chen, 2024). The **enabler** here is explicit linkage between activities and mechanisms; the corresponding **risk** is activity menus without theory, which dilute impact and hamper interpretation.

Another recurring design choice is **communal framing of STEM**. Framing STEM work as service to people and communities—while giving participants concrete opportunities to “act communally”—appears repeatedly, often supported by role models who narrate how their STEM work meets communal goals (Boucher et al, 2017; Prieto-Rodriguez et al., 2020; Beck, M. et al., 2021). This functions as an **enabler** when the communal purpose is made visible in tasks and stories, not merely in messaging.

With respect to **role-model strategy**, the corpus converges on **quality over mere exposure**. Role models are most influential when they are perceived as competent and attainable/similar, and when exposure is tailored rather than generic (Gladstone and Cimpian, 2021; Lawner et al., 2019; Prieto-Rodriguez et al., 2020). The **enabler** is curated, contextualised interaction; the **risk** is generic talks that do not translate into mechanism change.



Programmes also privilege **hands-on, authentic, sustained work**. Project-based learning, authentic design/build challenges, and collaborative problem-solving are repeatedly identified as productive design cores (Reinhold et al., 2018; Beck, M. et al., 2021; Torres-Torres et al., 2024; Prieto-Rodriguez et al., 2020). **Enablers** include multi-week projects with progressive challenge and coached reflection.

To sustain engagement over time, entries describe **hybrid formats and continuity**—combinations of after-school and in-school strands, visible next steps such as competitions, showcases, and internships, and progressive challenge across grades (Acuña et al., 2022; Dalela and Ahmed, 2024; Verdugo-Castro et al., 2022). The **enabler** is a clear, staged pathway; the **risk** is isolated, one-off experiences.

Finally, several entries emphasise **mentoring and networks**. Multi-mentor constellations, near-peer networks, and cohort designs (including living-learning communities) are highlighted as more effective than single, highly formal pairings, particularly for reducing isolation and improving persistence (Beck, M. et al., 2021; Verdugo-Castro et al., 2022). **Enablers** here are diversified mentoring and cohort continuity, while the **risk** is reliance on a single dyad without broader support. Complementing this, **teacher and interpersonal levers**—interactive engagement and facilitation that deliberately shape the interpersonal climate—are stressed; several entries note positive effects associated with interventions delivered by female teachers and with strategies that operate at the interpersonal environment level (Yu et al., 2024; Beck, M. et al., 2021). **Family stimulation and early orientation** also appear as supportive strands, with designs that involve parents and provide stimulation materials positioned as strengthening girls' views of science and supporting continuity (Avolio et al., 2020; Ha et al., 2023). Alongside learner-level work, **environmental/system supports**—visibility, supportive spaces, institutional signals—are presented as complements to classroom design (Yu et al., 2024; Beck, M. et al., 2021; Msambwa et al., 2024a).

Delivery components and patterns (how principles are enacted)

The reviewed entries pay close attention to enactment. **Structured collaboration and technical role rotation** translate inclusive intent into routine practice: rotating technical roles, explicit participation routines, and formative feedback that normalises struggle help sustain engagement and prevent dominance patterns (Beck, M. et al., 2021; Reinhold et al., 2018). These features are **enablers** when consistently applied.

Invention/engineering challenges and authentic production further anchor learning. Invention education, design-build challenges, and coached participation in competitions are associated with creative confidence, identity development, and skills—**when embedded in supportive coaching rather than framed as pure selection**² (Dalela and Ahmed, 2024; Reinhold et al., 2018; Torres-Torres et al., 2024). Here, supportive coaching is the **enabler**; selection framing is the **risk**.

² Treating competitions or challenges mainly as screening/ranking mechanisms—i.e., events that identify and reward those who already perform well—rather than as instructional experiences that build all participants' skills.



Where the intent is consolidation of confidence and identity, entries discuss **girls-only strands and stereotype-resistant content (where appropriate)**. Effects strengthen when such strands are combined with explicit stereotype-resistant content and progressive challenge (Acuña et al., 2022; Reinhold et al., 2018; Prieto-Rodríguez et al., 2020). **Cohorts** and **multi-mentor constellations** are connected to improvements in retention, learning outcomes, and perceived support, with diversified mentoring across seniority levels emphasised as an **enabler** (Beck, M. et al., 2021).

A number of entries present **growth-mindset plus experiential strands** as a practical complement to project-based designs (Beroíza-Valenzuela and Salas-Guzmán, 2024). In parallel, **field-image work and advising/pathways** ensure that the “who belongs” question and the “what comes next” question are addressed together; advising and navigable course maps make continuity visible and thereby **enable** progression (Wang and Yu, 2023; Plasman et al., 2021; Verdugo-Castro et al., 2022).

Measurement portfolios (what gets tracked, and with what evidence types)

Across entries, **core constructs** are repeatedly tracked: interest/attitudes, self-efficacy/confidence, identity/belonging, and markers of continuity such as course-taking, retention, and transition (Acuña et al., 2022; Beck, M. et al., 2021; Prieto-Rodríguez et al., 2020; Reinhold et al., 2018; Verdugo-Castro et al., 2022; Pérez-Felkner et al., 2025). Emotional responses to specific fields and orientation toward STEM careers are also noted (Acuña et al., 2022).

Evidence types and instruments most frequently include pre-post surveys, classroom/club observations, and interviews/focus groups, often complemented by administrative/academic indicators (e.g., GPA, graduation, awards, retention/continuation) and—where applicable—environmental outputs such as policy changes or dedicated spaces (Beck, M. et al., 2021; Prieto-Rodríguez et al., 2020; Reinhold et al., 2018; Pérez-Felkner et al., 2025). Several entries underscore **longitudinal follow-up and continuity indicators**, arguing that follow-up beyond immediate post-tests is required to capture sustained engagement and pathway uptake (Acuña et al., 2022; Prieto-Rodríguez et al., 2020). Programmes that create visible sequences tend to include **transition/retention markers** rather than attitudes alone (Plasman et al., 2021; Verdugo-Castro et al., 2022).

Illustrative portfolios show this range in practice. Beck, M. et al. (2021) combine retention and professional success with attitudes/motivation, leadership development, learning outcomes/GPA, promotions/awards, and environmental outputs (e.g., lactation rooms). Prieto-Rodríguez et al. (2020) report observations, pre-post surveys, and telephone/face-to-face interviews, noting that qualitative evidence frequently shows higher engagement than brief surveys. Reinhold et al. (2018) track pre-post attitudinal and orientation measures within girls-only project-based strands, with career mentoring/advisory elements linked to intentions. Pérez-Felkner et al. (2025) consolidate evaluation methodologies for inclusive environments and encourage diversified evidence collections matched to design. Acuña et al. (2022) point to frequent use of interest, self-efficacy, identity, and emotions



toward fields, and call for designs that document **how** components produce outcomes over time (not just **that** change occurred).

Design–measurement alignment (practical implications for programme planning)

The practical implications are consistent across entries.

First, **make the mechanisms measurable**: when programmes aim to build identity, belonging, and self-efficacy, indicators should reflect these constructs and be paired with behavioural continuity markers (Acuña et al., 2022; Prieto-Rodríguez et al., 2020; Verdugo-Castro et al., 2022; Beck, M. et al., 2021).

Second, **monitor enactment, not only outcomes**: because impact hinges on delivery quality, process indicators—participation balance, role rotation, feedback routines, mentoring contact—help interpret outcome movement (Beck, M. et al., 2021; Reinhold et al., 2018; Torres-Torres et al., 2024).

Third, **capture interpersonal and environmental levers** by including measures of teacher practices, classroom climate, visibility actions, and policy/space outputs alongside learner outcomes (Yu et al., 2024; Beck, M. et al., 2021; Msambwa et al., 2024a).

Finally, **embed continuity checks**: where programmes provide advising/maps, staged challenges, or progression opportunities, transition and retention should be tracked explicitly, not just near-term attitudes (Plasman et al., 2021; Verdugo-Castro et al., 2022; Boucher et al., 2017).

Conclusion

Taken together, the programme-level findings indicate that **explicit theory linkage, communal purpose, curated role-model interactions**, and **sustained, coached, hands-on work** are the core **enablers** when enacted through **structured collaboration** and supported by **hybrid pathways, mentoring networks, teacher-led climate work, family stimulation, and environmental signals**.

Evaluation practice that pairs **mechanism measures** with **continuity and implementation indicators** reduces ambiguity and supports iterative improvement. The principal **risks/barriers** arise when exposure is generic, activities are disconnected from mechanisms, opportunities are one-off, or the interpersonal and environmental context is not addressed.

Recommendations

This section presents actionable guidance derived from the reviewed evidence on programmes and factors influencing girls' participation in STEM. Recommendations are organised to align programme design, day-to-day delivery, and evaluation with the same underlying mechanisms of change. Each item specifies practical levers, associated **enablers**, and potential **risks** or barriers.



Mechanism-aligned measurement

Adopt a measurement plan in which *identity, belonging, self-efficacy, and interest* are specified as the core intermediate outcomes and are tracked alongside **behavioural continuity markers** (course-taking, transitions, retention). Employ **mixed methods** (pre–post surveys, observations, interviews/focus groups, artefacts) with **longitudinal follow-ups**, and document **implementation and environmental indicators** (participation balance, role rotation, feedback routines, mentoring contact; policy/space/visibility actions) to support interpretation of change. **Enablers** include the explicit pairing of mechanism measures with continuity and implementation data; **risks** arise when attitude measures are collected without delivery context or follow-up. (Acuña et al., 2022; Prieto-Rodríguez et al., 2020; Reinhold et al., 2018; Beck, M. et al., 2021; Verdugo-Castro et al., 2022; Pérez-Felkner et al., 2025; Yu et al., 2024)

Equity-centred programme stance

Position learner-level supports within a broader **institutional** approach that addresses bias and androcentric norms through **policy, space, and visibility** measures. Teacher development, inclusive spaces, and the visibility of women in STEM function as systemic complements to classroom practice. **Enablers** are coordinated actions across interpersonal and institutional layers; **barriers** persist where structural signals remain unchanged. (Beck, M. et al., 2021; Msambwa et al., 2024a; Msambwa et al., 2024b; Yu et al., 2024)

Designing the day-to-day experience

Make inclusive pedagogy observable in routine practice: **structured collaboration, rotation of technical roles, guided participation, and formative feedback** that normalises productive struggle. Provide **multi-week, hands-on, authentic projects** and coached invention/engineering challenges; make **pathways explicit** (advising, sequenced try-outs, showcases/competitions); and **engage families** from early stages. **Enablers** are regularised routines and visible progression; **risks** include one-off activities and unstructured group work. (Reinhold et al., 2018; Beck, M. et al., 2021; Dalela & Ahmed, 2024; Torres-Torres et al., 2024; Plasman et al., 2021; Verdugo-Castro et al., 2022; Avolio et al., 2020; Ha et al., 2023)

Role-model components as conditional

Treat role-model exposure as a **testable component** rather than a default. Where used, **curate for attainability/similarity and interaction quality**, embed encounters within **mentoring constellations** (near-peers and practitioners), and **instrument the mechanism** (perceived similarity, dosage, timing) with **comparative checks** against alternative uses of time (e.g., additional coached build time). If no incremental benefit is observed, **reallocate effort** to higher-yield components (inclusive pedagogy, collaboration architecture). **Enablers** are matched exemplars and measurable hypotheses; **risks** include generic talks and opportunity costs. (Lawner et al., 2019; Gladstone & Cimpian, 2021; Prieto-Rodríguez et al., 2020; Beck, M. et al., 2021; Verdugo-Castro et al., 2022; Bong & Chen, 2024)



Countering stereotypes and “brilliance” beliefs

Address stereotypes and “brilliance” beliefs through **multiple channels**. Embed counter-stereotypic cues within materials, spaces, facilitation, and narratives; make **communal purpose** salient in tasks and communications; and use **discipline-specific examples** to shift field images. **Enablers** are consistent, embedded practices; **risks** follow when efforts are limited to ad-hoc messaging or a single speaker event. (Boucher et al., 2017; Gladstone & Cimpian, 2021; Wang & Yu, 2023; Beck, M. et al., 2021)

Content priorities with inclusive pedagogy

Prioritise **computational thinking** and **invention education** as core literacies delivered through **feedback-rich, inclusive designs**. Adapt pedagogy to **disciplinary contexts** (e.g., physics) rather than relying on generic add-ons. **Enablers** are discipline-sensitive routines; **risks** arise when content is transplanted without pedagogical adaptation. (Torres-Torres et al., 2024; Dalela & Ahmed, 2024; Maries et al., 2024; Wang & Yu, 2023)

Planning for interactions and context

Plan for **interaction effects** and **contextual moderation**. Anticipate non-linear, site-dependent outcomes; design monitoring to detect **combined impacts** (e.g., role rotation + advising + stereotype-countering) and to account for **cultural and implementation differences**. **Enablers** include interaction-aware analysis and site-level context notes; **risks** occur when results are generalised without attention to moderation. (van Tuijl & van der Molen, 2016; Bong & Chen, 2024; Acuña et al., 2022; Wang & Yu, 2023; Maries et al., 2024)

Sustaining hybrid formats and continuity

Sustain **hybrid formats and continuity pathways** by blending after-school and in-school experiences, building **visible next steps** (internships, showcases), minimising early lock-in through **try-outs**, and tracking **transitions/retention** across stages. **Enablers** are clearly signposted routes and progressive challenge; **risks** include isolated experiences without bridges to subsequent opportunities. (Acuña et al., 2022; Verdugo-Castro et al., 2022; Plasman et al., 2021; Pérez-Felkner et al., 2025)

Conclusion

Collectively, the recommendations specify the mechanisms to target, the delivery practices to institutionalise, and the evidence to gather for appraisal over time. Implemented together, they support iterative improvement within sites and responsible adaptation across contexts.



Logic model and impact evaluation development

Logic model

The logic model formalises how HK Unicorn Squad is expected to create change for girls in STEM. It integrates (a) the review-level evidence on mechanisms and delivery conditions and (b) context-specific insights from the anthropological study of programme experience and perceived effects (Jaanits et al., 2025). Consistent with social-cognitive career theory and expectancy-value theory, the model positions **identity/belonging, self-efficacy, interest, and academic self-concept** as the primary intermediate outcomes through which persistent participation and progression are achieved (Prieto-Rodriguez et al., 2020; Reinhold et al., 2018; Beck, M. et al., 2021).

Activities are specified as **hands-on, multi-week projects** and invention/engineering strands; **structured collaboration** with rotating technical roles and formative feedback; **relevance and communal framing** of STEM work; **advising and try-before-specialising** opportunities; and complementary **teacher/interpersonal and environmental levers** (inclusive discourse, visibility of women, supportive spaces/policies). These features are the delivery conditions most consistently associated in the literature with gains in girls' interest, confidence, belonging, and continuity (Reinhold et al., 2018; Beck, M. et al., 2021; Torres-Torres et al., 2024; Boucher et al., 2017; Plasman et al., 2021; Verdugo-Castro et al., 2022; Yu et al., 2024). The anthropological material converges with these levers by documenting confidence to try, the salience of a psychologically safe climate, and the meaning children derive from collaborative making (Jaanits et al., 2025).

The model also recognises **conditional components**. Role-model exposure is included only where it is curated for attainability/similarity and embedded in learning, reflecting mixed empirical signals and the risk of limited effects from generic talks (Lawner et al., 2019; Gladstone & Cimpian, 2021; Prieto-Rodriguez et al., 2020). Finally, the model differentiates **assumptions already in place** (e.g., girls-only setting, early reach, manualised content) from **assumptions to strengthen** (e.g., explicit self-efficacy routines, parent touchpoints, pathway scaffolds), aligning with improvement priorities surfaced in both the literature and qualitative findings (Acuña et al., 2022; Pérez-Felkner et al., 2025; Jaanits et al., 2025).



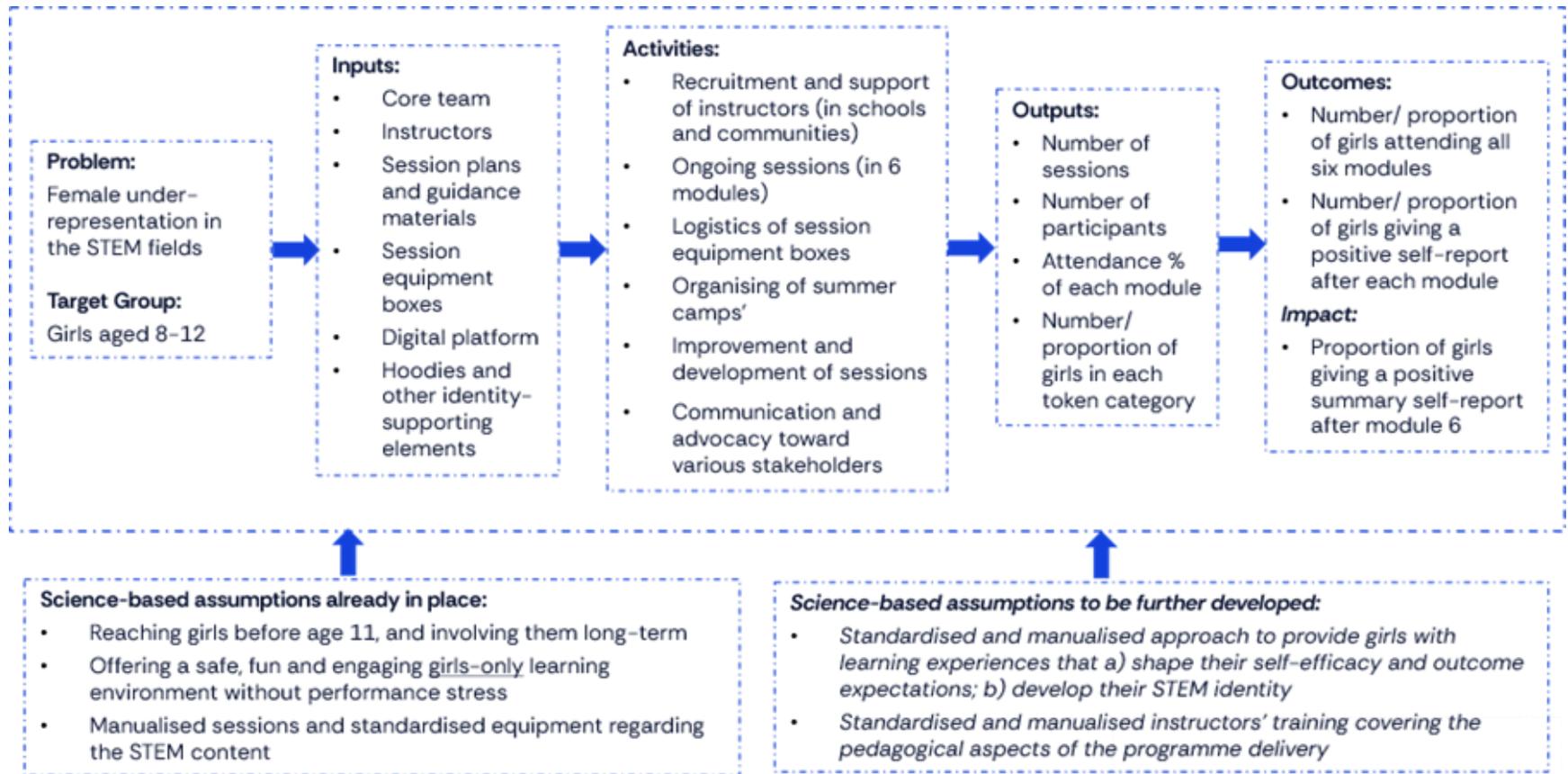


Figure 1. Logic model of the HK Unicorn Squad programme.



The diagram highlights a sequence linking **inputs** (people, manuals, equipment, partnerships) to **activities** (session sequences, camps, coached showcases, advising), **outputs** (dose, reach, participation balance, artefacts), and **short-/medium-/long-term outcomes**.

Emphasis is placed on two cross-cutting ideas. First, each activity strand is explicitly tied to one or more mechanisms (e.g., mastery → self-efficacy; communal framing → value/identity) to preserve theoretical clarity (Reinhold et al., 2018; Prieto-Rodriguez et al., 2020). Second, the strength of the arrows is **moderated by enactment quality**: rotating roles, feedback frequency, and inclusion prompts determine whether putative benefits materialise, consistent with evidence that format labels (club/course; online/in-person) are themselves neutral (Beck, M. et al., 2021; Bong & Chen, 2024).

Parallel rows for **teacher/interpersonal** and **environmental** levers acknowledge that system signals shape learner-level change (Yu et al., 2024; Msambwa et al., 2024a, 2024b). Dotted connections indicate hypothesised links to be verified through routine monitoring and follow-up (Plasman et al., 2021; Verdugo-Castro et al., 2022; Pérez-Felkner et al., 2025).

Science-based elements checklist

The checklist operationalises the logic model into **observable, rateable elements** that can be verified in delivery, standardised where appropriate, and tracked over time. Items are organised under timing/continuity, climate, learning design, progression supports, family/pathway touchpoints, outputs/reach, quick outcome signals, and instructor capability/support. Each item carries three status descriptors—*in place*, *manualised*, and *adoption*—to distinguish design availability from actual uptake.

The evidence base motivating the items is threefold.

1. **Learning design:** sustained project-based strands, coached production, and structured collaboration are associated with increases in interest, confidence, and identity (Reinhold et al., 2018; Beck, M. et al., 2021; Torres-Torres et al., 2024; Prieto-Rodriguez et al., 2020; Dalela & Ahmed, 2024).
2. **Climate:** identity-safe routines (equitable turn-taking, rotating technical roles, normalising struggle) counter dominance patterns and predict engagement and belonging (Beck, M. et al., 2021; Boucher et al., 2017; Wang & Yu, 2023).
3. **Progression:** advising, staged try-outs, and visible sequences are linked to transitions and retention (Plasman et al., 2021; Verdugo-Castro et al., 2022; Pérez-Felkner et al., 2025).

Role-model elements are presented **conditionally** and require checks for perceived similarity/attainability and interaction quality, reflecting the nuanced evidence on role-modeling (Lawner et al., 2019; Gladstone & Cimpian, 2021). In all domains, the anthropological study offers convergent validity by documenting the perceived value of



hands-on work, the importance of safe participation norms, and the conditions under which girls report growth in confidence (Jaanits et al., 2025).

Table 1. Science-based elements checklist of the HK Unicorn Squad programme.

Core element	Current level of achievement		
	In place (now)	Manualised	Adoption
A. Critical Timing & Transition Continuity			
A.1 Reach girls before age 11. [Interest often dips ~11–12; stereotypes can appear by ~6.]	Yes	Yes	Yes
A.2 Bridge the move to middle/secondary. [Keep sessions through this window to avoid drop-off.]	Partial	No	No
B. Identity-safe environment			
B.1 Girls-only where mixed groups disadvantage girls. [Use girls-only when mixed settings sideline girls or raise stereotype threat.]	Yes	Yes	Yes
B.2 Spaces/materials signal “me belonging in STEM”. [Remove masculine-coded décor; diverse girls/women in visuals/examples; inclusive language; tools sized for small hands.]	Yes	No	Yes
C. Duration & continuity			
C.1 Participation spans months/years (not days). [Longer exposure sustains interest and identity.]	Yes	Yes	Partial
C.2 Returning levels/tiers to sustain momentum. [e.g., L1 Foundations → L2 Integrated Projects → L3 Design Challenge / near-peer mentoring; visible progression (badges/micro-credentials) and “next module” invites.]	Partial	No	Partial
D. Activities that create learning experiences			
D.1 Mastery moments every session. [Each girl finishes a task independently.] (SCCT: ↑ self-efficacy; EVT: ↑ expectancy)	Partial	No	Yes



Core element	Current level of achievement		
	In place (now)	Manualised	Adoption
D.2 Near-peer models visible and similar. [Prioritise similar/attainable peers; if adults visit, ensure similarity + attainable path; evidence on “role models” is mixed, so emphasise near-peer similarity.] (SCCT: vicarious efficacy; EVT: ↑ attainment value)	Yes	No	Yes
D.3 Specific, credible feedback. [“Your wiring test worked because you isolated the fault.” Not generic praise.] (SCCT: ↑ efficacy; EVT: ↓ perceived cost)	No	No	Partial
D.4 Positive emotion & calm supported. [Short wins, choice, peer help; normalise “productive struggle.”] (EVT: ↑ intrinsic value; ↓ anxiety cost)	Yes	No	Yes
D.5 Spatial-strategy practice embedded. [Weekly micro-tasks tied to builds: mental rotation/visualisation from schematics, mirror/perspective tasks, paper-folding, block assembly, simple CAD (computer-aided design). <i>Why here? Reviews link spatial skills to maths and show they are trainable; practice is recommended.</i>]	Partial	Yes	Partial
E. Curriculum & pedagogy (identity-safe, meaningful, integrated)			
E.1 Hands-on tasks; equitable role rotation. [Rotate coder/builder/tester/presenter; no “note-taker only,” turn-taking group norms.]	Partial	No	Yes
E.2 Make value visible (utility + communal). [Link tasks to helping people/real problems; say how skills help school/next course/life.] (EVT: ↑ utility & attainment value)	No	No	No
E.3 Integrated STEM tasks (not fragmented). [Connect maths/CS/engineering/science within one project; LR notes integration yields more meaningful, real-world connections than siloed subjects.]	Partial	No	Partial
F. Family & pathway supports			



Core element	Current level of achievement		
	In place (now)	Manualised	Adoption
F.1 Parent touchpoints (i.e., systematic approach to guide each parent). [Brief encouragement prompts parents can send (10–20 words, effort/strategy-focused) + micro-infosheets (why this matters, how to ask about projects, next steps). <i>Literature review shows parent encouragement predicts girls’ persistence; simple scripts are a realistic way to shift daily conversations.</i>]	No	No	No
F.2 Support next-step choices with timely info. [In the Unicorn context: list next-term modules, local workshops/museums, summer programmes; how to request a school project/topic.]	Partial (not outside HKUS)	No	Partial (not outside HKUS)
G. Exposure & preparation			
G.1 Programming / CS exposure early. [CS = computer science; start with block-based coding, logic, debugging. <i>Why singled out? Literature review notes less prior CS experience among girls and that CS coursework exposure predicts interest & preparedness.</i>]	Yes	Yes	Yes
G.2 Math success opportunities (small wins + scaffolds). [Embed “math mini-wins” in builds (measurement, ratios, angles). Scaffolds = worked examples, visual aids, checklists, simple physical tools (rulers, blocks, paper models), chunked steps, quick feedback; celebrate correct reasoning, not speed.]	No	No	No
H. Outputs & reach (delivery basics)			
H.1 Attendance / session dose tracked. [Essential to interpret outcomes.]	Partial	Partial	Partial
H.2 Instructor:girl ratio set & monitored. [Quality guardrail so each girl gets mastery + feedback; set a local target that reliably allows this.]	Yes	No	Yes
I. Quick outcome signals			
I.1 Micro-questionnaires on interest / self-efficacy / belonging. [1–3 brief items after selected sessions.]	Partial	No	Partial



Core element	Current level of achievement		
	In place (now)	Manualised	Adoption
I.2 Track choices/persistence. [Where clubs/electives are limited: re-enrolment next term, attendance streaks, opting into harder modules, home STEM time, borrowing kits, signing up for external workshops/competitions.]	No (not outside HKUS)	No	No
J. Instructor capability & support			
J.1 Session Content Manuals (STEM-specific) – materials exist & usable. Objectives; key concepts; misconceptions; step-by-step method (with pictures); materials/safety; troubleshooting; success criteria. Supports accurate content delivery and consistent mastery opportunities.]	Yes	Yes	Yes
J.2 STEM content knowledge & subject didactics by instructors. [Instructors know the STEM concepts they teach, typical misconceptions, and use subject-specific didactics: e.g., multiple representations, error analysis, stepwise progression. Why? Strong content + pedagogical content knowledge supports reliable mastery experiences and accurate feedback, building self-efficacy.]	Partial	Partial	Partial
J.3 Session Pedagogical Checklist. [E.g., open/close routine; role rotation; turn-taking norms; formative checks to run; feedback script; time boxes; inclusion cues.]	No	No	No
J.4 Instructor Facilitation & Inclusive Pedagogy Micro-skills. [E.g., runs role rotation & turn-taking; uses formative checks; gives specific feedback; applies chunking; builds in retrieval and spiral practice; uses think-pair-share.]	Partial	No	No
J.5 Peer observation / quick coaching loop. [E.g., occasional 10-minute peer drop-ins; 2 warm suggestions + 1 tweak.]	Partial	No	Partial

Interpretation of the checklist presented in Table 1 proceeds along three lines. First, gaps where elements are *in place* but not *manualised* indicate standardisation opportunities; where *manualised* but low *adoption*, the implication is targeted coaching, simplification, or reinforcement through routine prompts (Beck, M. et al., 2021; Yu et al., 2024). Second, prioritisation favours **mechanism-proximal and feasible** changes—e.g., mastery feedback routines, participation balance, and relevance prompts—before layering broader pathway and



family supports (Acuña et al., 2022; Plasman et al., 2021; Jaanits et al., 2025). Third, conditional components (role-modeling) should be accompanied by data on similarity/attainability and compared with alternative uses of time (Lawner et al., 2019; Gladstone & Cimpian, 2021). Annual re-rating then functions as a light-touch continuous-improvement cycle, linked to the logic model's outcome signals (Verdugo-Castro et al., 2022; Pérez-Felkner et al., 2025).

Post-module questionnaire

The questionnaire provides routine, minimal-burden evidence aligned to the logic model's intermediate outcomes and delivery conditions. Designed for a few minutes on phones/tablets, it uses child-friendly response formats while preserving construct intent. Item families reflect the literature-based mechanisms and the experiences reported in the anthropological study.

- **Self-efficacy/mastery & productive struggle:** items capturing perceived capability, task success, and persistence when challenged (Reinhold et al., 2018; Prieto-Rodriguez et al., 2020; Beck, M. et al., 2021; Jaanits et al., 2025).
- **Belonging/identity & climate:** items on feeling safe to contribute and being heard, consistent with the importance of inclusive interactional norms (Beck, M. et al., 2021; Boucher et al., 2017; Wang & Yu, 2023; Jaanits et al., 2025).
- **Interest/value & communal purpose:** items assessing enjoyment/utility and the sense that STEM can help people/communities (Boucher et al., 2017; Acuña et al., 2022; Torres-Torres et al., 2024).
- **Transfer & continuation intention:** items on trying activities at home/school and willingness to return or progress (Verdugo-Castro et al., 2022; Plasman et al., 2021; Pérez-Felkner et al., 2025).
- **Role-model checks (only when relevant):** items on perceived similarity/attainability and novel learning from the encounter, given the contingent effects in the evidence base (Lawner et al., 2019; Gladstone & Cimpian, 2021).

Table 2. Suggested post-module questionnaire of the HK Unicorn Squad programme.

These short questions help us improve Unicorn Squad. There are no right or wrong answers. Please choose the face that best matches what you think after your participation.

Response options (5-point)

- 😄 Strongly agree
- 😊 Agree
- 😐 Neither agree nor disagree
- 😞 Disagree
- 😡 Strongly disagree

Replace [MODULE TOPIC] with the main focus of the module.



1. I found the [MODULE TOPIC] very interesting. (*interest and enjoyment*)
2. I dare to try new things related to the [MODULE TOPIC]. (*confidence*)
3. I can now solve challenges related to the [MODULE TOPIC] that would previously have been too difficult for me. (*mastery development / self-efficacy*)
4. This was a safe place to experiment with different things, ask questions, and make mistakes. (*safe climate*)
5. Even when it was difficult at times, I still wanted to continue so that I could reach a successful result. (*persistence and perceived "cost" relative to outcome*)
6. I liked working in a group with other girls, and I always had a chance to share my ideas. (*collaboration and communication*)
7. I can show or teach at home the things we tried here. (*transfer of knowledge and skills / usefulness*)
8. I can show or teach at school the things we tried here. (*transfer of knowledge and skills / usefulness*)
9. I feel that I fully belong to the Unicorn Squad community. (*identity / belonging*)
10. I can imagine continuing with the [MAIN TOPIC OF THE MODULE] in the future—for example, studying it further or working in this field. (*identity and future choices*)

The instrument is designed to be age-appropriate and brief. For younger respondents, the **emoji-anchored five-point scale** reduces reading load and supports stable responding; for older cohorts, the same items can be administered with a standard Likert scale without icons. Retaining a **neutral option** ("Neither agree nor disagree") is recommended to minimise forced choices and improve reliability, particularly when respondents are unsure or have mixed experiences.

The current item order—interest, self-efficacy, persistence, mastery cue, climate, participation/voice, utility/transfer, belonging, and future intention—tracks the typical **SCCT/EVT** causal sequence without implying a performance test. This ordering helps avoid priming effects that could arise from beginning with evaluative climate items or long-term intentions.

When a role-model encounter takes place in a given session, two optional add-on items can be included to test incremental contribution: **perceived similarity/attainability** and **identity projection** ("I can imagine myself doing this kind of work"). These items operationalise SCCT vicarious experience and outcome expectations. They should be analysed separately from the



core mechanism set and, where feasible, **compared to sessions or sites without role-model exposure** to estimate added value rather than assuming it.

For scoring, responses should be coded **5 → 1 (Strongly agree → Strongly disagree)**. At this early stage, avoid reverse-scored items to reduce respondent error and simplify interpretation; if evidence of acquiescence bias³ emerges later, a small number of reverse-worded items can be piloted with cognitive interviewing. Site-level means and distributions should be reviewed routinely, with particular attention to **mechanism–practice links**: for example, low scores on the participation/voice item indicate strengthening role rotation and explicit participation routines, while low climate scores indicate a need to reinforce inclusion prompts and **error-positive feedback**.

Interpretation should connect proximal signals to delivery evidence and to **behavioural continuity**. Intention to continue (the final item) ought to be tracked against **actual continuity indicators** (subsequent enrolment, retention, transitions) to validate predictive utility over time. **Aggregation** by module and by site is recommended to detect trends and to attribute changes to specific delivery adjustments documented in the checklist (e.g., coaching on feedback routines, adoption of role rotation). In this way, the questionnaire functions as a **lightweight, mechanism-aligned monitor** within a broader **SCCT/EVT-informed improvement cycle**.

Findings of the questionnaire are read against the logic model and checklist to guide iterative improvement. Persistently lower scores on **belonging** or **productive struggle** indicate a need to reinforce inclusive routines (e.g., role rotation, feedback practices) that the literature associates with better outcomes (Beck, M. et al., 2021; Reinhold et al., 2018; Bong & Chen, 2024). Increases in **continuation intention** should co-occur with **transitions/retention**; if not, additional pathway scaffolds—advising checkpoints, visible next steps—are warranted (Plasman et al., 2021; Verdugo–Castro et al., 2022; Pérez–Felkner et al., 2025). Where **role-model** items are used, results should be compared with alternative allocations of session time (e.g., additional coached building). Lack of added value implies reallocating minutes to higher-yield routines, consistent with the contingent status of role-modeling in the evidence (Lawner et al., 2019; Gladstone & Cimpian, 2021). Patterns by site/age cohort are examined to identify contextual moderators; in line with the evidence on system signals, improvements may require teacher practice adjustments or changes to visibility/space in addition to learner-level tweaks (Yu et al., 2024; Msambwa et al., 2024a, 2024b; Jaanits et al., 2025).

³ The respondent’s tendency to agree with statements regardless of their content—especially when items are positively worded, the topic feels evaluative, or power dynamics (adult–child, teacher–student) make “agree” feel safe.



References

- Acuña, S. A., Michaelis, J. E., Roth, J. D., & Towles, J. (2022). Interventions to increase young people's interest in STEM: A scoping review. *Frontiers in Psychology, 13*, 954996. <https://doi.org/10.3389/fpsyg.2022.954996>
- Avolio, B., Chávez, J., & Vílchez-Román, C. (2020). Factors that contribute to the under-representation of women in science careers worldwide: A literature review. *Social Psychology of Education, 23*, 773–794. <https://doi.org/10.1007/s11218-020-09558-y>
- Beck, M., Cadwell, J., Kern, A., Wu, K., Dickerson, M., & Howard, M. (2022). Critical feminist analysis of STEM mentoring programmes: A meta-synthesis of the existing literature. *Gender, Work & Organisation, 29*①, 167–187. <https://doi.org/10.1111/gwao.12729>
- Beroíza-Valenzuela, F., & Salas-Guzmán, N. (2024). STEM and gender gap: A systematic review in WoS, Scopus and ERIC databases (2012–2022). *Frontiers in Education, 9*, 1378640. <https://doi.org/10.3389/feduc.2024.1378640>
- Bong, D. C. N., & Chen, C. J. (2024). Insights into science, technology, engineering and mathematics (STEM) achievement: A comprehensive review of factors and methodologies. *Review of Education*. Advance online publication. <https://doi.org/10.1002/rev3.70018>
- Boucher, K. L., Fuesting, M. A., Diekman, A. B., & Murphy, M. C. (2017). Can I work with and help others in this field? How communal goals influence interest and participation in STEM fields. *Frontiers in Psychology, 8*, 901. <https://doi.org/10.3389/fpsyg.2017.00901>
- Dalela, S., & Ahmed, M. S. (2024). Systematic review of the invention education research landscape. *Frontiers in Education, 9*, 1284442. <https://doi.org/10.3389/feduc.2024.1284442>
- Gladstone, J. R., & Cimpian, A. (2021). Which role models are effective for which students? A systematic review and four recommendations for maximising effectiveness in STEM. *International Journal of STEM Education, 8*, Article 59. <https://doi.org/10.1186/s40594-021-00315-x>
- Ha, V. T., Hai, B. M., Duong, T. T. M., Nguyen, V. H., et al. (2023). Preschool STEM activities and associated outcomes: A scoping review. *International Journal of Engineering Pedagogy, 13*②, 100–116. <https://doi.org/10.3991/ijep.v13i8.42177>
- Jaanits, Jaanika, Sepp, Merylyn, Rajaveer, Liina, Lepik, Liina Maria (2025). "HK Unicorn Squad annabki selle võimaluse, et ta on valmis kastist välja mõtlema." HK Unicorn Squadi programmi kogemuse ja mõju uuring. Antropoloogia Keskus.
- Lawner, E. K., Quinn, D. M., Camacho, G., Johnson, B. T., & Pan-Weisz, B. (2019). Ingroup role models and underrepresented students' performance and interest in STEM: A meta-analysis of



lab and field studies. *Social Psychology of Education*, 22②, 1169–1195.
<https://doi.org/10.1007/s11218-019-09518-1>

Maries, A., Li, Y., & Singh, C. (2024). Challenges faced by women and persons excluded due to ethnicity and race in physics learning environments: A literature review. *Reports on Progress in Physics*, 88, 015901. <https://doi.org/10.1088/1361-6633/ad91c4>

Msambwa, M. M., Daniel, K., Cai, L., & Fute, A. (2024). A systematic review of the factors affecting girls' participation in STEM subjects. *Computer Applications in Engineering Education*, 32②, e22707. <https://doi.org/10.1002/cae.22707>

Msambwa, M. M., Daniel, K., Lianyu, C., & Antony, F. (2024). A systematic review using feminist perspectives on factors affecting girls' participation in STEM. *Science & Education*. Advance online publication. <https://doi.org/10.1007/s11191-024-00524-0>

Perez-Felkner, L., Erichsen, K., Li, Y., Chen, J., Hu, S., Ramirez Surmeier, L., & Shore, C. (2025). Computing education interventions to increase gender equity (2000–2020): A systematic review. *Review of Educational Research*, 95③, 536–580.
<https://doi.org/10.3102/00346543241241536>

Plasman, J., Gottfried, M., Williams, D., Ippolito, M., & Owens, A. (2021). Parents' occupations and students' success in STEM: A systematic review. *Adolescent Research Review*, 6, 33–44.
<https://doi.org/10.1007/s40894-020-00136-z>

Prieto-Rodríguez, E., Sincock, K., & Blackmore, K. (2020). STEM initiatives matter: Results from a systematic review of secondary-school interventions for girls. *International Journal of Science Education*, 42⑦, 1144–1161. <https://doi.org/10.1080/09500693.2020.1749909>

Reinhold, S., Holzberger, D., & Seidel, T. (2018). Encouraging a career in science: Secondary-school effects on student STEM orientation. *Studies in Science Education*, 54④, 69–103. <https://doi.org/10.1080/03057267.2018.1442900>

Torres-Torres, Y.-D., Román-González, M., & Pérez-González, J.-C. (2024). Didactic strategies for computational-thinking education from a gender perspective: A systematic review. *European Journal of Education*, 59②, e12640. <https://doi.org/10.1111/ejed.12640>

van Tuijl, C., & van der Molen, J. H. W. (2016). Study choice and career development in STEM: An overview and integration. *International Journal of Technology and Design Education*, 26, 159–183. <https://doi.org/10.1007/s10798-015-9308-1>

Verdugo-Castro, S., García-Holgado, A., & Sánchez-Gómez, M. C. (2022). The gender gap in higher STEM studies: A systematic review. *Heliyon*, 8②, e10300.
<https://doi.org/10.1016/j.heliyon.2022.e10300>



Wang, L. H., & Yu, Z. G. (2023). Gender-moderated effects of academic self-concept on achievement, motivation, performance and self-efficacy: A systematic review. *Frontiers in Psychology*, 14, 1136141. <https://doi.org/10.3389/fpsyg.2023.1136141>

Yu, W., He, J., Luo, J., & Shu, X. (2024). Interventions for gender equality in STEM education: A meta-analysis. *Journal of Computer Assisted Learning*, 40(6), 2558–2573. <https://doi.org/10.1111/jcal.12928>

