

26 **Abstract**

27 **Background**

28 Despite decades of calls for increased diversity in the health research workforce, disparities
29 exist for many populations, including Black, Indigenous, and People of Color individuals, those
30 from low-income families, and first-generation college students. To increase representation of
31 historically marginalized populations, there is a critical need to develop programs that
32 strengthen their path toward health research careers. High school is a critically important time to
33 catalyze interest and rebuild engagement among youth who may have previously felt excluded
34 from science, technology, engineering, and mathematics (STEM) and health research careers.

35 **Methods**

36 The overall objective of the MYHealth program is to engage high school students in a
37 community-based participatory research program focused on adolescent health. Investigators
38 will work alongside community partners to recruit 9th through 12th graders who self-identify as a
39 member of a group underrepresented in STEM or health research careers (e.g., based on race
40 and ethnicity, socioeconomic status, first generation college student, disability, etc.). MYHealth
41 students are trained to be co-researchers who work alongside academic researchers, which will
42 help them to envision themselves as scientists capable of positively impacting their communities
43 through research. Implemented in three phases, the MYHealth program aims to foster a
44 continuing interest in health research careers by developing: 1) researcher identities, 2)
45 scientific literacy, 3) scientific self-efficacy, and 4) teamwork and leadership self-efficacy. In
46 each phase, students will build knowledge and skills in research, ethics, data collection, data
47 analysis, and dissemination. Students will directly collaborate with and be mentored by a team
48 that includes investigators, community advisors, scientific advisors, and youth peers.

49 **Discussion**

50 Each year, a new cohort of up to 70 high school students will be enrolled in MYHealth. We
51 anticipate the MYHealth program will increase interest and persistence in STEM and health
52 research among groups that have been historically excluded in health research careers.

53 *Keywords:* children and adolescents, education, research, science, technology,
54 engineering & math (STEM), health disparities, community-based participatory research
55 (CBPR), identity, health sciences research

56

57

58

59

60

61

62

63

64

65

66

67 Introduction

68 Despite decades of calls for increased diversity in the research workforce, several groups
69 continue to be excluded and thereby underrepresented in Science, Technology, Engineering,
70 and Math (STEM) careers in the United States [1]. In 2019, Black, Indigenous, and People of
71 Color (BIPOC) individuals were awarded only 11.7% of STEM research doctorates, despite
72 making up about a third of the U.S. population and the workforce [1]. Likewise, while women
73 comprised 47% of the workforce, they received only 42.2% of STEM doctorates [1]. Parental
74 education is also associated with postsecondary enrollment, which is an important step on the
75 trajectory toward many STEM professions. College enrollment directly after high school is lower
76 among children whose parents did not graduate high school (50%), attend college (57%), or
77 graduate college (67%), compared to those whose parents have completed a bachelor's degree
78 (84%) [2]. Low-income youth also continue to graduate high school and enroll in college at lower
79 rates (65% vs. 83%) than their middle- and high-income peers [2]. While these figures have
80 improved slightly in the last decade, significant work is still needed to address these inequities
81 [3, 4].

82 An initial and persistent interest in science is driven largely by social and environmental
83 factors, making those from historically excluded groups more vulnerable to barriers such as
84 family factors (e.g., attitudes toward science careers, parent interest in scientific activities),
85 pedagogical factors (e.g., teaching culture, classroom culture), extracurricular activities (e.g.,
86 science clubs, out-of-school experiences), and social factors (e.g., models of science interest,
87 peer culture), which in turn impact one's expectations and beliefs that they can succeed in
88 science [5, 6]. High school and adolescence are critical windows to intervene in developing an
89 interest in STEM. This period of the life course is centered on identity development and is when
90 youth begin to understand their skills, interests, and values and begin to make career decisions
91 [6]. When youth are encouraged to practice science and strengthen their beliefs about their

92 efficacy, they are more likely to express interest, achieve, and persist in science. However, this
93 is not the most likely experience for youth from historically excluded groups, who may
94 experience school practices that do not nurture science interests and capacities [2, 7] and have
95 access to fewer out-of-school science experiences, mentorship, and science role models [2, 5].

96 Despite school requirements for science curricula, some high school students still lack
97 meaningful and relevant opportunities to engage in science and research [8, 9]. Research
98 suggests that teens from historically excluded groups often view school science as “hard” and
99 “discouraging” and do not see science as relevant to their everyday lives [5, 9-11]. Moreover,
100 students may not know any science professionals in their daily lives, making science and
101 research seem abstract. When paired with a lack of advocates, mentors, and support systems,
102 this leaves students from historically excluded groups with more negative experiences of
103 science [11]. On the other hand, when students experience success with science, receive
104 support from parents and other adults, and develop self-efficacy in science, they are more likely
105 to have positive attitudes toward science [8]. Out-of-school experiences, including summer
106 research experiences and after-school programs can help students develop an interest in
107 science, improve science self-efficacy, increase their science knowledge and skills, and build
108 commitment and persistence in high school that may improve enrollment and retention in
109 undergraduate and graduate school programs in the future [12-14]. Critical to the success and
110 effectiveness of programs is exposure to developmentally- and personally-relevant research
111 opportunities, mentorship and social support, and experiences that encourage empowerment,
112 agency, and a sense of purpose [9].

113 Increasing the participation and success of researchers from historically excluded
114 groups is essential for strengthening innovation and global leadership in health research [15].
115 Diverse and inclusive scientific teams not only generate novel research questions and methods,
116 but by bringing their life experiences and perspectives, they also offer new problem-solving

117 approaches to persistent issues in research [16]. Furthermore, tackling the innumerable health
118 disparities faced by historically excluded groups requires the support of diverse scientific teams.
119 Thus, the need for a diverse health research workforce serves as a comprehensive theoretical
120 and empirical foundation for the Michigan Youth Health (MYHealth) program [6, 11, 12, 17]. The
121 overall goal of the MYHealth research training program is to address the persistent problem of
122 underrepresentation of historically excluded groups in health research careers. This project
123 combines an innovative, youth-centered research program with a community-based
124 participatory research (CBPR) approach to partner with high school students as co-researchers.
125 The training program aims to help youth build persistence toward health research careers by
126 developing: 1) researcher identities, 2) scientific literacy, 3) scientific self-efficacy, and 4)
127 teamwork and leadership self-efficacy. Engaging high school students in a research training
128 program is advantageous because postsecondary decisions and occupational interests are
129 often crystallized and plans formulated during adolescence [9].

130 **Conceptual overview**

131 Because students from excluded groups are more likely to encounter external barriers (e.g.,
132 fewer STEM role models, less access to rigorous STEM coursework) that are associated with
133 dropping out of the STEM and health research pipeline, it is imperative to understand how to
134 support their persistence toward science careers effectively [18-20]. Based on an ethnically
135 diverse sample, Chemers and colleagues demonstrated that researcher identity, science self-
136 efficacy, and leadership and teamwork self-efficacy mediated the relationship between science
137 support experiences and persistence in science and health research careers [18]. Extrapolating
138 from these findings, researchers have validated the effectiveness of interventions that
139 incorporate a range of out-of-school programs that provide opportunities for research
140 experiences, mentoring, and community involvement [12-14]. Building on Chemers' framework
141 for understanding persistence in science, MYHealth integrates social cognitive theory,

142 expectancy-value theory of achievement motivation, and CBPR. The MYHealth framework is
143 outlined in Fig 1.

144 **Fig 1. MYHealth conceptual framework, adapted from Chemers et al., 2011**

145 **Social cognitive theory (SCT)**

146 SCT seeks to explain human learning and motivation and has been applied to domain-specific
147 learning (e.g., science learning) across fields [21-25]. SCT identifies three components that
148 interact to influence learning: personal characteristics (e.g., self-efficacy), environmental
149 characteristics (e.g., school, support, community), and behavior (e.g., pursuit of opportunities).
150 These dynamics have been extensively studied in relation to science interest, achievement, and
151 career choice [18, 26, 27], and among historically excluded youth [28, 29]. Research
152 demonstrates that students from historically excluded groups have fewer opportunities to have
153 their science self-efficacy beliefs nurtured; these dynamics, in turn, suppress motivation and
154 science self-efficacy beliefs and ultimately contribute to fewer youth pursuing science degrees
155 and research careers [30-32].

156 **Expectancy-value theory of achievement motivation**

157 Linked to SCT, the core of the expectancy-value theory of achievement motivation is that
158 students' achievement and decisions related to achievement are determined by expectancies
159 for success in that domain and the perceived value of that domain in their lives [33, 34].
160 According to this theory, achieving and maintaining an interest in science results from students'
161 beliefs about their likelihood of success in science and the value they assign to science. Applied
162 to MYHealth, expectancy-value theory posits that students from excluded groups can be
163 supported to view science and research careers as attainable and important through
164 opportunities that: 1) increase positive perceptions related to their science and research skills,

165 characteristics, and competencies and 2) increase positive perceptions related to their personal
166 values and goals. Taken together, science- and research-related SCT and expectancy-value
167 theory help us to understand the processes that contribute to persistence in science and
168 research and associated careers, particularly among historically excluded students.

169 **Community-based participatory research (CBPR)**

170 CBPR is built on core principles that support partnership, capacity building, co-learning, and
171 action toward equity and social justice [26, 35, 36] and has decades of research demonstrating
172 its effectiveness in creating partnerships and enhancing the quality and rigor of research [24,
173 37-41]. CBPR follows the research cycle and includes building a research team (with
174 nontraditional research partners, like youth), identifying the needs and priorities of a community,
175 asking questions, gathering data, interpreting data from the lens of the community, and
176 implementing action strategies toward health and social justice.

177 Using CBPR as a model for engaging youth as co-researchers benefits youth directly
178 (e.g., research training, communication skills, problem-solving and critical thinking skills,
179 leadership, preparedness for college), their communities (e.g., community literacy, improved
180 programs and services, improved policy), and research quality (e.g., improved recruitment,
181 retention, sustainability) [42, 43]. Given its overall inclusive and equitable approach, CBPR is a
182 compelling and effective methodology for engaging students from marginalized groups and
183 supporting their long-term interest in science.

184 In summary, the MYHealth program builds on the theoretical underpinnings of SCT,
185 expectancy-value theory of achievement motivation, and a CBPR approach to create an out-of-
186 school research experience for youth that includes experiences with an applied adolescent
187 health research project, instrumental and socioemotional mentoring from academic researchers
188 and a community of learners, and community involvement through research projects that impact

189 youth and their communities. Through these program components, we will impact four primary
190 psychological and cognitive processes: researcher identity, science self-efficacy, leadership and
191 teamwork self-efficacy, and scientific literacy. By using MYHealth to develop these domains
192 among youth co-researchers, we will achieve our long-term goal of increasing persistence
193 toward health research careers.

194 **Overview of the MYHealth program**

195 The MYHealth program is a research training program with the long-term goal of increasing the
196 representation of groups traditionally excluded from biomedical, behavioral, and clinical
197 research, including individuals from low-income backgrounds, first-generation college students,
198 and racial and ethnic minority youth. The MYHealth program will enroll 4 successive 12-month
199 cohorts of high school students who self-identify as underrepresented in STEM. The MYHealth
200 program achieves its overall aim via multiple phases each year, as outlined in Fig 2. The
201 Summer Launch, Impact Projects, and Peer Leadership phases each have specific goals,
202 activities, outputs, and outcomes that are summarized in Table 1 and detailed in the sections
203 that follow. The program curriculum was developed through an iterative process of piloting and
204 feedback from high school students from the local population. A condensed curriculum was
205 presented to a pilot cohort of participants and the refined curriculum is detailed below. Each
206 year, modifications will be made based on our formative evaluation.

207 **Fig 2. MYHealth program overview**

208 **Table 1. MYHealth program structure**

Phase <i>[Timeline]</i>	Activities	Outputs	Outcomes
-----------------------------------	-------------------	----------------	-----------------

Summer Launch <i>[July/August]</i>	Health sciences research training	Materials on website Social media presence	Interest in science careers Researcher identity, science self-efficacy, scientific literacy
Impact Projects <i>[September – May]</i>	Bi-monthly research meetings Study design Data collection Data analysis Impact projects	Original research Conference on Adolescent Health presentation Impact Project (e.g., posters, social media campaigns)	Advanced research training and application Researcher identity, science self-efficacy, and scientific literacy Understanding of impact through dissemination Recognition of the real-world relevance and translation of research
Peer Leadership <i>[August – May]</i>	Lead and support research activities	Meeting logs Social media presence	Teamwork and leadership self-efficacy Persistence in science
Evaluation <i>[Ongoing]</i>	Evaluation and iterative improvement	Lessons learned Publications	Model of research training and capacity-building for youth

209 **Summer launch**

210 The Summer Launch will be a 10-day training in CBPR and health research methodologies for
 211 up to 70 students each year. The overall goal is to train high school students to be co-
 212 researchers who can work alongside academic researchers on studies of adolescent health to
 213 develop their interest in STEM and incubate researcher identities. A sample of Summer Launch
 214 curriculum topics and program constructs are outlined in Table 2. The Summer Launch is
 215 designed to take place in person on the University of Michigan campus to build excitement and
 216 energy related to research, create community among students from different communities in
 217 Southeast Michigan, and introduce youth to college campuses. The Summer Launch will be

218 adapted and conducted virtually when meeting in person is not possible due to COVID-19
 219 restrictions.

220 Throughout Summer Launch, students will learn about research methods, CBPR, and
 221 various adolescent health studies (Table 2). We will create applied learning experiences using
 222 MyVoice, a national text message poll of 14-24-year-olds to understand adolescents’ views on a
 223 range of health-related topics [44]. For all discussions, MyVoice will serve as a specific data
 224 source to illustrate research elements in the context of a real-world project. For example, when
 225 learning about data collection, youth will review existing MyVoice question sets and participate
 226 in a MyVoice pilot, receiving SMS messages like MyVoice participants. Consistent with our
 227 CBPR approach that trains youth to be co-researchers, the Summer Launch agenda and
 228 activities will align with equity, inclusion, and social justice principles. Each topic will consider
 229 the importance of ethical and responsible conduct of research—from designing a study with the
 230 well-being and needs of participants in mind, to obtaining consent, to fairly representing data
 231 and disseminating it to all relevant stakeholders.

232 **Table 2. Sample summer launch agenda topics aligned with MYHealth constructs**

MYHealth Constructs		
Researcher Identity	Science Self-Efficacy	Leadership and Teamwork Self-Efficacy
Teambuilding	Iterative Research Cycles	Teambuilding Staying Involved: Opportunities with MYHealth Research to Action Teambuilding and Planning
Introduction to MYHealth	Doing Ethical Treatment in Research	
MyVoice Simulation: Building a Study	Research with People, Not on People	
Who is a Researcher?	Designing MYHealth	
What’s In It For You, Your Community, and For Health?	What Counts as Data?	
Careers in Research: Panel	Qualitative Data Collection	

of Researchers	Quantitative Data Collection	
Tailoring Messages Based on Audience	Participatory Data Collection	
What Happens Next?	Text Messaging for Data Collection	
Impact of MYHealth: Sharing Findings to Make Change	What is Analysis?	
Staying Involved: Opportunities with MYHealth	Teaching the World What You Learned	
	Research to Action Teambuilding and Planning	

233 **Impact projects**

234 At the end of the Summer Launch, all students are invited to apply to participate in the Impact
 235 Projects phase, a nine-month program that aims to build on the overall momentum, skills, and
 236 interests developed in the previous phase. When applying, students will specify their interest in
 237 various adolescent health topics to help guide the direction of the Impact Projects. We anticipate
 238 that up to 30 high school students will complete Impact Projects in the following academic year.

239 Impact Projects are guided by several key principles, including equitable partnership as
 240 co-researchers, positioning youth as experts of their own experiences, studying a relevant and
 241 meaningful topic to youth, and sharing findings with diverse audiences. By engaging in these
 242 collaborative activities, the students will develop their researcher identities, enhance their sense
 243 of scientific self-efficacy, and hone their scientific literacy.

244 Students will meet virtually, approximately twice per month, in small teams that include a
 245 faculty mentor, an undergraduate and/or graduate student research assistant, and up to two
 246 Peer Leaders who are MYHealth alumni. Projects are conducted virtually to broaden access to
 247 MYHealth for high school students across schools and communities. To effectively engage high
 248 school students, meetings will be highly interactive, including group processes and applied

249 exercises. See Table 3 for a sample of the Impact Projects agenda. Activities will mimic the
 250 research process, including three key phases: 1) Study Design; 2) Data Collection and Analysis;
 251 and 3) Dissemination. These phases will include ongoing training in CBPR, health research
 252 methods, and the responsible conduct of research. Though most of the training and activities
 253 will occur in small groups, each team will also share their progress with other Impact Project
 254 teams and program faculty to receive feedback and practice research communication.

255 **Table 3. Impact projects agenda topics by stage**

Impact Project Stages		
<i>[Timeline]</i>		
Study Design	Study Design	Study Design
<i>[September – October]</i>	<i>[September – October]</i>	<i>[September – October]</i>
Identifying a priority area	Identifying a priority area	Identifying a priority area
Gathering and summarizing existing literature	Gathering and summarizing existing literature	Gathering and summarizing existing literature
Developing a research question	Developing a research question	Developing a research question
Writing a research plan	Writing a research plan	Writing a research plan
Developing, piloting, and revising open-ended survey questions	Developing, piloting, and revising open-ended survey questions	Developing, piloting, and revising open-ended survey questions

256 **Peer leadership**

257 The overall aim of the Peer Leadership phase is to provide ongoing social and academic
 258 support, increase motivation, and improve self-efficacy for students from historically excluded
 259 groups pursuing STEM and health research careers [28, 41]. Peer Leaders will receive
 260 additional coaching and support from program faculty, including additional training in CBPR and
 261 health research methodologies. Selection criteria for being a Peer Leader include having an
 262 aptitude and interest in becoming a researcher. We anticipate that up to 10 of the 30

263 participating Impact Project students each year will be Peer Leaders in the following year.

264 Peer Leaders will interact closely with MYHealth high school students and program
265 faculty during the Summer Launch and Impact Projects. Peer Leaders will be part of Impact
266 Project teams and will provide coaching and guidance to their team members, work alongside
267 faculty mentors, and maintain logs of research activities and progress for the program
268 evaluation. Peer Leaders will be encouraged to contribute to dissemination efforts, including 1)
269 contributing to manuscripts and presentations to share evaluation findings; 2) supporting
270 resource-sharing efforts through in-person presentations to local STEM programs; 3) supporting
271 resource-sharing efforts through the website and social media strategies.

272 **Methods**

273 A mixed methods approach will be used to collect multiple quantitative and qualitative measures
274 of implementation and outcomes over five years. An explanatory sequential design (quantitative
275 followed by qualitative) will be implemented wherein survey responses will inform final interview
276 questions to evaluate the impact of the MYHealth program on participants' interest in and
277 persistence in STEM research. The Institutional Review Board at the University of Michigan
278 approved the evaluation study of the MYHealth program (HUM00213914, HUM00214694,
279 HUM00214949). All MYHealth participants will consent or assent to participate in the research
280 and parental consent will be collected for minors. Participants will be assigned a random
281 identifier at the start of the program to protect their anonymity and confidentiality throughout
282 data analysis. Because survey questions are focused on attitudes and beliefs about science and
283 research and are not sensitive in nature, there is minimal risk to the participants throughout the
284 course of the study.

285 **Setting**

286 The MYHealth program focuses on recruiting high school students based in Southeast Michigan
287 including, but not limited to, the Detroit, Southfield, Flint, East Lansing, Ypsilanti, and the
288 surrounding areas. These communities in Southeastern Michigan will be well-suited for
289 recruitment of high school students from historically excluded groups and have a clear need for
290 additional science training that is more engaging, relevant, and impactful than traditional
291 classroom science. During the 2016-2017 school year, science proficiency rates in these
292 communities lagged behind other schools in Southeastern Michigan and the state of Michigan.
293 For example, several focal communities had science proficiency rates of 11th grade high school
294 students between 9% and 11%. When looking at economically disadvantaged districts, these
295 numbers dropped below 8%. While state proficiency assessments are not the sole determinant
296 of science interest, enrollment in science courses in college, or future STEM careers [45], high
297 school science achievement can alter youth's perceptions of their capacity to engage in science
298 [6, 18, 46].

299 **Recruitment**

300 High school students will be recruited through ongoing partnerships with community partners
301 who have established networks of schools, teachers, and communities and will connect eligible
302 students and families to the MYHealth program. These partners include university centers that
303 serve students and schools, local school districts, and an academic support program. Our
304 community partners will distribute recruitment materials directly (e.g., information sheets, flyers,
305 emails) through their networks. Interested students and their families will be directed to
306 complete an online application form or contact our research team via email or telephone. The
307 application form (administered via the Qualtrics secure platform) asks interested high school
308 students and their families to provide contact information, state their interest in MYHealth, and
309 explain why they think they would be a good fit for the program. Strong applicants will

310 demonstrate an interest in health, community engagement, research, or the capacity to work in
311 teams.

312 **Eligibility criteria**

313 Incoming high school students living in Southeast Michigan who self-identify as members of an
314 underrepresented group in science or health will be invited to participate. No racial or ethnic
315 groups will be excluded from participating in the study.

316 **Data collection and outcome measures**

317 Numerous scales that have been previously validated with youth will be used to measure youth
318 experiences with STEM and research [8], researcher identity [38, 40], interest in and
319 persistence in STEM careers [24], scientific literacy [37, 39], and leadership, and teamwork-self-
320 efficacy [18]. Measures were tested during a pilot phase and then revisions were incorporated.
321 Table 4 presents an overview of the key measures and the critical time points at which they are
322 collected.

323 **Surveys**

324 An online survey will be administered online via Qualtrics at multiple timepoints during each
325 program phase, including pre- and post- Summer Launch, pre-, mid-, and post- Impact Projects,
326 pre- and post- Peer Leadership, and one year after completing all program activities (Table 4).
327 The one-year follow-up survey will track youths' continued interest and persistence in STEM
328 research careers, knowledge and skills related to STEM research, and how the program has
329 influenced their engagement in other research activities.

330 **Engagement and participation**

331 Each year students' reactions to the Summer Launch and the Impact Project programs will be
332 collected to make ongoing course improvements. In addition, we will design Peer Leader and
333 graduate research assistant logs to facilitate reflections about what worked in each meeting and
334 what could be improved before the next session for both the Summer Launch and the Impact
335 Project programs.

336 A research assistant will track program hours and attendance to gauge dosage [47] in
337 the Impact Projects. Logs will be used to track the implementation of CBPR processes at
338 monthly meetings. The logs will consist of a *Research Process Checklist* [48] to understand
339 progress in the research design process (e.g., stated research problem, defined variables, and
340 specified procedures to analyze data). Finally, the graduate research assistant will complete an
341 observational rating scale, *Youth Participatory Action Research Process Template* [49], to track
342 CBPR processes demonstrated by the Peer Leader and Impact Project students in each small
343 group, including training and practice of research skills, promoting strategic thinking, group
344 work, networking, communication skills, and power sharing. In addition, Evaluators will work
345 with the graduate research assistants to complete the *Critical Thinking Rubric for PBL* [50] at
346 three points during the academic year. In the spring of each year, Impact Project students will
347 disseminate research findings. The graduate research assistant will use the *Presentation Rubric*
348 *for PBL* to capture how students are communicating research processes and findings with
349 youth, community, policy, and academic audiences [51].

350 **Semi-structured interviews and/or focus groups**

351 Each year at the end of the Summer Launch, up to 10 students will be invited to participate in a
352 semi-structured interview or focus group [52]. We will use a semi-structured interview guide that
353 explores student experiences during the Summer Launch, including perceptions of research and

354 STEM careers, experiences with diversity, equity, and inclusion during MYHealth, and possible
 355 barriers to participation and continued engagement. Interviews will be recorded and transcribed.

356 **Table 4. Overview of evaluation instruments and data collection timepoints**

Measure	Outcomes and Definition	Administration
Surveys		
Participant Demographics & “Is Science Me?” Questionnaire [8]	Age, gender, race/ethnicity, disability status, household characteristics, previous experience with STEM	Pre-Summer Launch Pre-Impact Projects Pre-, Post-Peer Leadership tenure
Researcher Identity Scale [41, 42]	Youth researcher identity (extent to which youth see themselves as researchers)	Pre-, Post-Summer Launch Pre-, Mid-, Post-Impact Projects Post-Peer Leadership tenure
Science Motivation Questionnaire II (SMQ-II; [43])	Interest and persistence in STEM research careers (youth knowledge, attitudes and interest in STEM research careers)	Pre-, Post-Summer Launch Pre-, Mid-, Post-Impact Projects Post-Peer Leadership tenure
Research Self-Efficacy [53]	Youth research self-efficacy (confidence applying and using science concepts)	Pre-, Mid-, Post-Impact Projects Post-Peer Leadership tenure One-Year follow-up
Critical Thinking Rubric for Problem-Based Learning (PBL), adapted to assess skills in data analysis and reporting [47]	Youth scientific literacy (application of science practices and crosscutting concepts)	Pre-, Mid-, Post-Impact Projects
Presentation Rubric for PBL [48]		
Scientific Literacy Assessment [44, 45]		
Leadership and Teamwork Self-Efficacy [18]	Leadership and teamwork self-efficacy (confidence leading and working on a research team)	Pre-, Post-Peer Leadership tenure

Engagement and Participation

*Impact Project Logs [46]	Engagement with MYHealth program (hours & attendance)	Following monthly Impact Project meetings
*Research Process Checklist [47]	Participatory research skills (knowledge and application of research skills and CBPR processes)	Following monthly Impact Project meetings
*Youth Participatory Action Research (YPAR) Process Template [48]		

Semi-Structured Interview/Focus Group

Internally developed interview/focus group guide	Student experiences including perceptions of research and STEM careers, experiences with diversity, equity, and inclusion during MYHealth and possible barriers to participation and continued engagement	Post-Summer Launch Post-Impact Projects Post-Peer Leadership tenure
--	---	---

357 *instrument completed by research assistant

358 Similarly, at the end of the Impact Project each year, a sample of up to 10 students will
 359 be asked to participate in semi-structured interviews or a focus group to understand processes
 360 that occurred during implementation with each cohort. Questions will probe topics such as
 361 students’ perceptions of STEM and health research careers, their experiences with diversity,
 362 equity, and inclusion on the Impact Projects, and possible barriers to participation and continued
 363 engagement. All interviews will be recorded and transcribed for analysis.

364 **Analyses**

365 Overall, there are parallel strategies to evaluate all three program components: the Summer
 366 Launch, Impact Projects, and Peer Leadership. A Repeated Measures Analysis of Variance will
 367 be conducted to analyze the quantitative items in the pre-, post-, and follow-up surveys. Small-
 368 scale analyses will be performed on each program phase, and a longer-term evaluation will be
 369 conducted on longitudinal data. Over time, between-subjects factors may be added to account
 370 for differences between cohorts, or disaggregate data by demographic factors (e.g., gender,
 371 ethnicity). The final reporting will utilize appropriate statistical significance and associated effect

372 sizes.

373 Interview data will be analyzed with an inductive Outcome Harvesting approach [54, 55],
374 in which stakeholders identify individual and group impacts that mattered most, and speculate
375 on program elements that led to these outcomes. Codes will be applied to all interview
376 transcripts and then synthesized into related themes and conclusions. Findings will be used to
377 inform future years of the MYHealth program.

378 **Results**

379 We anticipate that up to 280 total high school students from historically excluded groups will be
380 enrolled in MYHealth across four years: up to 280 will have completed the Summer Launch, a
381 subset of up to 120 will have completed the Impact Projects, and a smaller subset of up to 40
382 will have become Peer Leaders. Primary quantitative outcomes for this evaluation are
383 Researcher Identity and STEM research career interest. Our power analyses indicate that we
384 need 90 students each in the Summer Launch and Impact Projects to detect a minimum effect
385 size of 0.3 which is considered small by conventional interpretations [56]. This assumes 80%
386 power at 5% Type 1 error (two-tailed t-test). Given our anticipated sample, we will have
387 sufficient students to make conclusions about changes in Researcher Identity and STEM career
388 interest even with attrition.

389 For the qualitative components of the program evaluation, we will ensure that we have a
390 sufficient sample size to achieve thematic saturation. In qualitative approaches, thematic
391 saturation refers to the point at which there is adequate evidence to develop themes and no
392 new insights are gathered with continued data collection. For interview studies, thematic
393 saturation is often estimated to be achieved between 6-20 interviews [57, 58].

394

395 **Discussion**

396 MYHealth is limited by our small cohort size and geography. The cohort size for each phase is
397 limited to ensure that students receive adequate time with faculty mentors and support for their
398 small group projects. Over the four years of the project, we will be able to understand the impact
399 of the research training program. Currently, only students from Southeast Michigan can
400 participate. If successful, future iterations of MYHealth could leverage our virtual program to
401 extend to another setting outside of Southeast Michigan.

402 The first complete year of the MYHealth program began in the Summer of 2022, and will
403 re-open annually for the following four years. We anticipate data will become available through
404 peer-reviewed publications beginning in the first two years and on several more instances
405 throughout the course of the project.

406 **Conclusion**

407 The MYHealth program capitalizes on Social Cognitive Theory, Expectancy Value Theory, and
408 a CBPR approach to design and implement a program that aims to address the continuing lack
409 of diversity within the STEM workforce. The program assesses the role and intersection of
410 personal characteristics (e.g., researcher identity, self-efficacy), social and environmental
411 factors, and behavioral influences to spark and maintain interest in STEM careers.

412 **Authors' contributions**

413 All authors, MD, JR, AA,TC, MDi, KH, LMV, DCW contributed equally to the the work by
414 conceptualizing the study aims and writing, reviewing and editing the study protocol.

415 **References**

- 416 1. National Center for Science and Engineering Statistics (NCSES) [Internet]. Women,
417 Minorities, and Persons with Disabilities in Science and Engineering. Directorate for
418 Social, Behavioral and Economic Sciences. Alexandria, VA: National Science
419 Foundation; 2021 [cited 2022 May 5]. Report No.: Special Report NSF-21-321:[Available
420 from: <https://nces.nsf.gov/pubs/nsf21321/>.
- 421 2. National Center for Science and Engineering Statistics (NCSES) [Internet]. Science and
422 Engineering Indicators. National Science Board. Alexandria, VA: National Science
423 Foundation; 2021 [cited 2022 May 5]. Available from:
424 <https://nces.nsf.gov/pubs/nsb20196/data>.
- 425 3. NIH research: Action needed to ensure workforce diversity strategic goals are achieved.
426 Bethesda (MD): US Government Accountability Office (GAO); 2018. Report No.: GAO-
427 18-545.
- 428 4. Fry R, Kennedy B, Funk C. STEM jobs see uneven progress in increasing gender, racial
429 and ethnic diversity. Washington (DC): Pew Research Center Science & Society; 2021.
- 430 5. Barton AC, Tan E, Rivet A. Creating hybrid spaces for engaging school science among
431 urban middle school girls. *Am Educ Res J*. 2008;45(1):68-103.
- 432 6. Britner SL, Pajares F. Sources of science self-efficacy beliefs of middle school students.
433 *J Res Sci Teach*. 2006;43(5):485-99.
- 434 7. Vedder-Weiss D, Fortus D. Adolescents' declining motivation to learn science: Inevitable
435 or not? *Res Sci Teach*. 2011;48(2):199-216.
- 436 8. Aschbacher PR, Li E, Roth EJ. Is science me? High school students' identities,
437 participation and aspirations in science, engineering, and medicine. *J Res Sci Teach*.
438 2010;47(5):564-82.

- 439 9. Barton AC, Tan E. Funds of knowledge and discourses and hybrid space. *J Res Sci*
440 *Teach.* 2009;46(1):50-73.
- 441 10. Kang H, Calabrese Barton A, Tan E, D. Simpkins S, Rhee H-y, Turner C. How do middle
442 school girls of color develop STEM identities? Middle school girls' participation in
443 science activities and identification with STEM careers. *Sci Educ.* 2019;103(2):418-39.
- 444 11. Zacharia Z, Barton AC. Urban middle-school students' attitudes toward a defined
445 science. *Sci Educ.* 2004;88(2):197-222.
- 446 12. Dirks C, Cunningham M. Enhancing diversity in science: Is teaching science process
447 skills the answer? *CBE—Life Sci Educ.* 2006;5:218-26.
- 448 13. McLean NA, Fraser M, Primus NA, Joseph MA. Introducing students of color to health
449 sciences research: An evaluation of the health disparities summer internship program. *J*
450 *Community Health.* 2018;43(5):1002-10.
- 451 14. Salto LM, Rigg ML, De Leon DD, Casiano CA, De Leon M. Underrepresented minority
452 high school and college students report STEM-pipeline sustaining gains after
453 participating in the Loma Linda University summer health disparities research program.
454 *PLoS One.* 2014;9(9).
- 455 15. National Academy of Sciences, National Academy of Engineering, and Institute of
456 Medicine. *Expanding Underrepresented Minority Participation: America's Science and*
457 *Technology Talent at the Crossroads.* Washington (DC): The National Academies Press;
458 2011.
- 459 16. Swartz TH, Palermo AS, Masur SK, Aberg JA. The science and value of diversity:
460 Closing the gaps in our understanding of inclusion and diversity. *J Infect Dis.*
461 2019;220(Suppl 2):s33-s41.
- 462 17. Andersen L, Ward TJ. Expectancy-value models for the STEM persistence plans of
463 ninth-grade, high-ability students: A comparison between Black, Hispanic, and White
464 students. *Sci Educ.* 2014;98(2):216-42.

- 465 18. Chemers MM, Zurbriggen EL, Syed M, Goza BK, Bearman S. The role of efficacy and
466 identity in science career commitment among underrepresented minority students. *J*
467 *Soc Issues*. 2011;67(3):469-91.
- 468 19. Hazari Z, Sadler PM, Sonnert G. The science identity of college students: Exploring the
469 intersection of gender, race, and ethnicity. *J Coll Sci Teach*. 2013;42(5):82-91.
- 470 20. Hazari Z, Sonnert G, Sadler PM, Shanahan M-C. Connecting high school physics
471 experiences, outcome expectations, physics identity, and physics career choice: A
472 gender study. *J Res Sci Teach*. 2010;47(8):978-1003.
- 473 21. Bandura A. *Social Foundations of Thought and Action: A Social Cognitive Theory*.
474 Englewood Cliffs (NJ): Prentice-Hall; 1986.
- 475 22. Bandura A. Human agency in social cognitive theory. *Am Psychol*. 1989;44(9):1175-84.
- 476 23. Bandura A. Social cognitive theory: An agentic perspective. *Annu Rev Psychol*.
477 2001;52(1):1-26.
- 478 24. Bryan RR, Glynn SM, Kittleson JM. Motivation, achievement, and advanced placement
479 intent of high school students learning science: Motivation of high school students
480 learning science. *Sci Educ*. 2011;95(6):1049-65.
- 481 25. Wang X. Why students choose STEM majors: Motivation, high school learning, and
482 postsecondary context of support. *Am Educ Res J*. 2013;50(5):1081-121.
- 483 26. Minkler M, Wallerstein N. *Community-based participatory research for health: From
484 process to outcomes*. 2nd ed. Hoboken (NJ): Jossey-Bass; 2010.
- 485 27. van der Zee T, Reich J. Open education science. *AERA Open*.
486 2018;4(3):2332858418787466.
- 487 28. Hurtado S, Cabrera NL, Lin MH, Arellano L, Espinosa LL. Diversifying science:
488 Underrepresented student experiences in structured research programs. *Res High Educ*
489 *J*. 2009;50:189-214.

- 490 29. Trujillo G, Tanner KD. Considering the role of affect in learning: monitoring students'
491 self-efficacy, sense of belonging, and science identity. *CBE Life Sci Educ.* 2014;13(1):6-
492 15.
- 493 30. Byars-Winston A, Estrada Y, Howard C, Davis D, Zalapa J. Influence of social cognitive
494 and ethnic variables on academic goals of underrepresented students in science and
495 engineering: A multiple-groups analysis. *J Couns Psychol.* 2010;57(2).
- 496 31. Fouad NA, Santana MC. SCCT and underrepresented populations in STEM fields:
497 Moving the needle. *J Career Assess.* 2017;25(1):24-39.
- 498 32. Navarro RL, Flores LY, Worthington RL. Mexican American middle school students' goal
499 intentions in mathematics and science: A test of social cognitive career theory. *J Couns*
500 *Psychol.* 2007;54(3).
- 501 33. Eccles J. Achievement and Achievement Motives: Psychological and Sociological
502 Approaches. In: Spence JT, editor. *Expectancies, Values, and Academic Behaviors.* San
503 Francisco (CA): Freeman; 1983. p. 75-146.
- 504 34. Wigfield A, Eccles JS. Expectancy–value theory of achievement motivation. *Contemp*
505 *Educ Psychol.* 2000;25(1):68-81.
- 506 35. Israel BA. *Methods in Community-Based Participatory Research for Health.* San
507 Francisco (CA): Jossey-Bass; 2005.
- 508 36. Jacquez F, Vaughn LM, Wagner E. Youth as partners, participants or passive recipients:
509 A review of children and adolescents in Community-Based Participatory Research
510 (CBPR). *Am J Community Psychol.* 2013;51:176-89.
- 511 37. Arnold ME, Bourdeau VD, Nott BD. Measuring science inquiry skills in youth
512 development programs: The science process skills inventory. *J Youth Dev.* 2018;8(1):15.
- 513 38. Bathia S, Morell L, Smith R, Koo B, Wilson M. Measuring high school students' self-
514 perception as a researcher: Development and validation of the researcher identity scale.

- 515 American Educational Research Association (AERA) Annual Meeting; San Francisco
516 (CA). 2020.
- 517 39. Bourdeau V, Arnold ME. Science Process Skills Inventory [Internet]. Corvallis (OR): 4-H
518 Youth Development Education, Oregon State University; 2009 [cited 2022 5 May].
519 Available from:
520 [http://www.pearweb.org/atis/data/documents/000/000/010/Science_Process_Skills_Inve](http://www.pearweb.org/atis/data/documents/000/000/010/Science_Process_Skills_Inventory_1.pdf)
521 [ntory_1.pdf](http://www.pearweb.org/atis/data/documents/000/000/010/Science_Process_Skills_Inventory_1.pdf).
- 522 40. Morell L, Bathia S, Koo B, Smith R, Wilson M. The construction and validation of a
523 research identity scale. American Educational Research Association (AERA) Annual
524 Meeting; San Francisco (CA). 2019.
- 525 41. Palmer RT, Maramba DC, Dancy TE. A qualitative investigation of factors promoting the
526 retention and persistence of students of color in STEM. *J Negro Educ.* 2011;80(4):491-
527 504.
- 528 42. DeJonckheere M, Vaughn LM, Bruck D. Youth-led participatory action research: A
529 collaborative methodology for health, education, and social change. *Sage Res Meth*
530 *Cases.* 2017.
- 531 43. Suleiman AB, Soleimanpour S, London J. Youth action for health through youth-led
532 research. *J Community Pr.* 2006;14(1-2):125-45.
- 533 44. DeJonckheere M, Nichols LP, Moniz MH, Sonnevillie KR, Vydiswaran VV, Zhao X, et al.
534 MyVoice national text message survey of youth aged 14 to 24 years: Study protocol.
535 *JMIR Res Protoc.* 2017;6(12):e8502.
- 536 45. Maltese AV, Tai RH. Pipeline persistence: Examining the association of educational
537 experiences with earned degrees in STEM among U.S. students. *Sci Edu.*
538 2011;95(5):877-907.
- 539 46. Chen JA, Usher EL. Profiles of the sources of science self-efficacy. *Learning and*
540 *Individual Differences.* 2013;24:11-21.

- 541 47. O'Donnell L, O'Donnell C, Wardlaw DM, Stueve A. Risk and resiliency factors
542 influencing suicidality among urban African American and Latino youth. *Am J*
543 *Community Psychol.* 2004;33(1-2):37-49.
- 544 48. Freed MN, Ryan JM, Hess RK. *Handbook of Statistical Procedures and Their Computer*
545 *Applications to Education and the Behavioral Sciences.* United Kingdom: Macmillan
546 Publishing Co, Inc; 1991.
- 547 49. Ozer EJ, Douglas L. Assessing the key processes of youth-led participatory research:
548 Psychometric analysis and application of an observational rating scale. *Youth Soc.*
549 2015;47(1):29-50.
- 550 50. PBLWorks [Internet]. 6-12 Critical Thinking Rubric (non-CCSS) Novato (CA). 2020 [cited
551 2022 5 May]. Available from:
552 https://my.pblworks.org/resource/document/6_12_critical_thinking_rubric_non_ccss.
- 553 51. PBLWorks [Internet]. 9-12 Presentation Rubric for PBL (non-CCSS Novato (CA). 2020
554 [cited 2022 5 May]. Available from:
555 [https://my.pblworks.org/system/files/documents/PBLWorks-9-12-Presentation-Rubric-](https://my.pblworks.org/system/files/documents/PBLWorks-9-12-Presentation-Rubric-Non-CCSS.pdf)
556 [Non-CCSS.pdf](https://my.pblworks.org/system/files/documents/PBLWorks-9-12-Presentation-Rubric-Non-CCSS.pdf).
- 557 52. DeJonckheere M, Vaughn LM. Semistructured interviewing in primary care research: A
558 balance of relationship and rigour. *FMCH.* 2019;7(2).
- 559 53. Ozer EJ, Schotland M. Psychological empowerment among urban youth: Measure
560 development and relationship to psychosocial functioning. *Health Edu & Behav.*
561 2011;38(4):348-56.
- 562 54. Abboud R, Claussen C. The use of Outcome Harvesting in learning-oriented and
563 collaborative inquiry approaches to evaluation: An example from Calgary, Alberta. *Eval*
564 *Program Plann.* 2016;59:47-54.
- 565 55. Wilson-Grau R. Outcome harvesting: Better Evaluation; 2015 [cited 2022 5 May].
566 Available from: https://www.betterevaluation.org/en/plan/approach/outcome_harvesting.

- 567 56. Maher JM, Markey JC, Ebert-May D. The other half of the story: Effect size analysis in
568 quantitative research. *CBE Life Sci Educ.* 2013;12(3):345-51.
- 569 57. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with
570 data saturation and variability. *Field Methods.* 2006;18(1):59-82.
- 571 58. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies:
572 Guided by information power. *Qual Health Res.* 2016;26(13):1753-60.
- 573

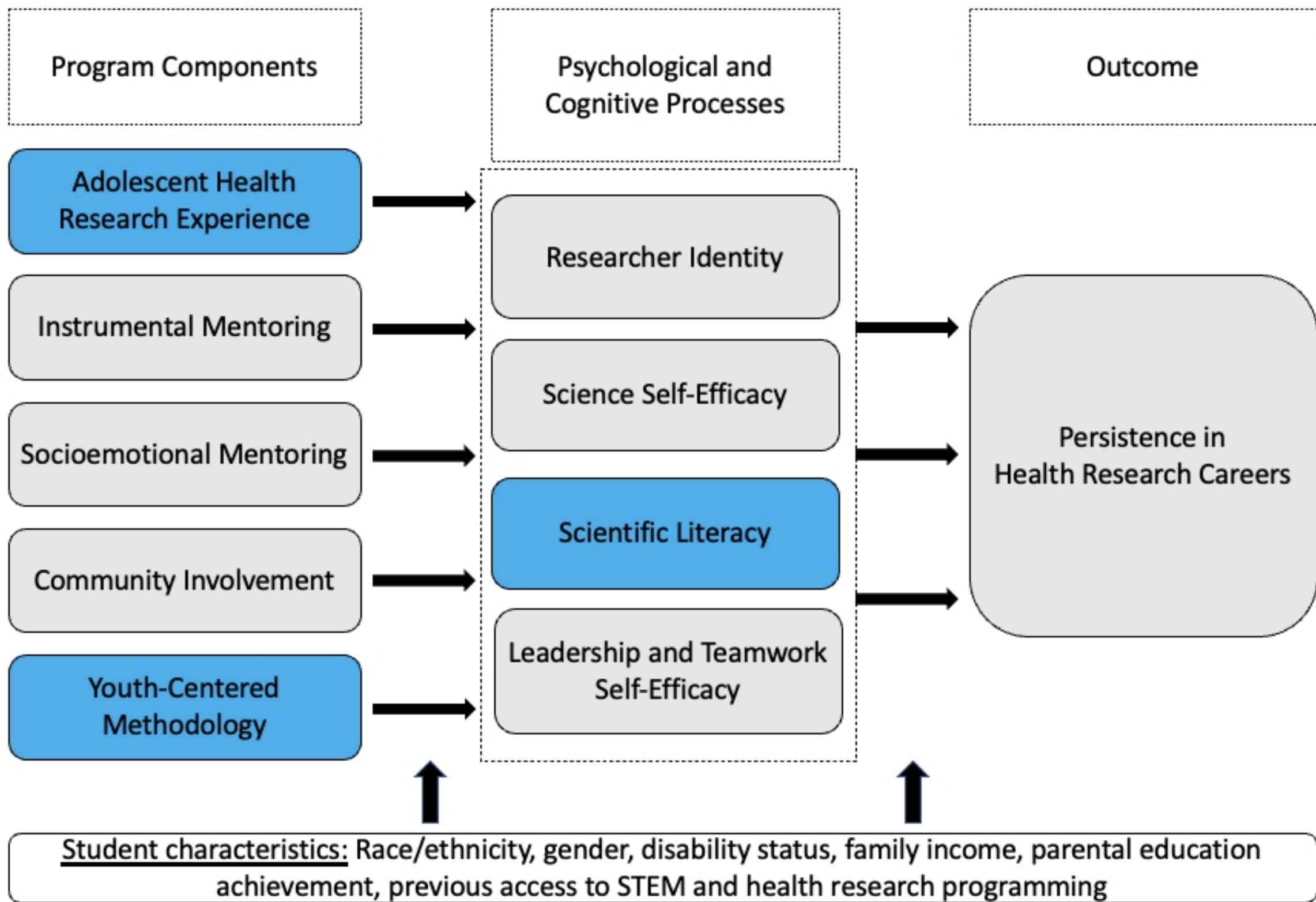


Fig1

1. Summer Launch

10-day training program highlighting health services research, CBPR, and increasing STEM persistence

2. Impact Projects

9-month program with an emphasis on further development and application of skills. Students complete a real-life research project in small teams throughout the academic year



3. Peer Leadership

MYHealth alumni return to gain additional training and mentorship from MYHealth faculty and participate in Summer Launch and Impact Project teams

Fig2