Broadening Participation Research Project: STEP into STEM,
Successful Transitions and Effective Pathways into STEM

1 Project Description

1.1 Vision Statement

STEP into STEM is a data-driven study that will (1) advance the knowledge and evidence base for successful strategies to encourage underrepresented minorities (URMs) at the high-school, community-college and 4-year-undergraduate level, who are either uncommitted or non-STEM majors, to commit to STEM pathways; and (2) develop a set of nationally-applicable policy recommendations (including measurements/metrics) that will dramatically increase the number of URMs that join and remain on STEM pathways to the STEM workforce.

1.1.1 Goals

STEP into STEM will study longitudinal data from high school through postsecondary education and beyond, will use various theoretical models that provide the framework for success in STEM undergraduate studies, and then will develop strategies and approaches that could be implemented to prepare HBCU undergraduate students for successful pursuit of graduate programs and/or careers in STEM. STEP into STEM will address emerging challenges/opportunities and the limited progress being made in improving broadening participation (BP) in STEM (Committee, 2013).

STEP into STEM goals include (1) characterizing and understanding pathways to STEM, by comparing and contrasting *alternative pathways* to more traditional *academic pipelines*; (2) assessing STEM-intervention effectiveness at universities across the nation based on institutional system/program implementations and metrics related to student success (STEM degree attainment); and (3) making recommendations to increase the number/percentage of URM/women populations in postsecondary institutions that enter STEM pathways (especially at HBCUs/MSIs where underrepresentation in STEM careers is influenced by complex equity, socioeconomic, and inclusion-related challenges).

STEP into STEM will investigate behavioral, cognitive, affective, learning and social differences as well as institutional/systematic/programmatic processes that impact BP in STEM. Ultimately, the goal is to export/disseminate the intellectual merit of this study nationally, and to provide data-driven approaches that produce successful pathways for URMs to the STEM workforce. The scale of potential improvement (numbers of students) that will benefit from alternative pathways to STEM will be estimated, to address the shortage of "deep analytical skills." (McKinsey Global Institute, 2011.)

1.1.2 Broadening Participation Challenges Addressed

A diverse, well-prepared, and innovative workforce and STEM-literate citizenry are crucial to the Nation's health and economy (President's Council, 2012). The number of URMs in STEM pathways is insufficient to address underrepresentation in both colleges and the STEM workforce (Hunt, 2015; NSF, 2015; Committee, 2014). In 2011, 26% of STEM workers were women, 9% were minorities (non-White, non-Asian) (NSF, 2011). Women and URMs are known to opt out of STEM majors at disproportionate rates (Xie, 2003). Bold, large-scale, national initiatives from K-20 are needed (NSF, 2014). Billions of federal dollars have been spent to attract students to STEM fields (GAO, 2012).

This study will provide knowledge that will address the enduring national challenge, and the STEM divide (Hopewell, 2009), to increase the number of URM students on a competitive STEM trajectory in high schools (HSs), community colleges (CCs), and 4-year undergraduate institutions.

1.1.3 Focus of Research

Extensive study of the STEM "pipeline" has taken place over a few decades. The hypothesis of the "pipeline model" is that if one increases the number/percentage of students entering the pipeline near the *source of the pipe* (that includes middle-school ages), then even with attrition/leakage that causes students to transition out of the STEM pipeline, there will be a significant rise in STEM graduates from college to the STEM workforce (Fealing, 2012). Evidence indicates that this approach is somewhat effective, but the demand for STEM workers is outpacing the STEM-prepared graduate pool (Snyder, 2011; and Staklis, 2010). This is particularly true for URMs in STEM. Many HSs, CCs, and 4-year universities/colleges have implemented programs and strategies to try to prevent the loss of students from the pipeline, with mixed success. The present research will evaluate the effectiveness of the STEM interventions that have been tried across the nation, but will most importantly look specifically at the transitions into STEM. *This research will focus on "Why students select and excel in STEM studies," rather than "Why students drop out."*

Many CC and 4-year incoming students enter without specifying a major (are either undecided or undeclared), or enter non-STEM majors at the beginning of their postsecondary education. This focused research effort will analyze these students, and will characterize and understand the driving forces that transition those students (to enter or re-enter STEM pipeline) from being initially undecided or in non-STEM majors to ultimately graduate with a STEM degree. One of the hypotheses is that there are alternative pathways (with effective support systems and programs) that undecided/non-STEM students can follow to achieve STEM degrees, and ultimately enter the STEM workforce.

1.1.4 Uniqueness and Innovation

This research uniquely examines the critical factors drawing students into STEM, particularly from undecided/non-STEM incoming students at the CC and 4-year college levels. Data from an NCES (National Center for Education Statistics) BPS:04/09 report (NCES, 2012) suggests significant numbers of students transition from non-STEM majors (social/behavior sciences, humanities, business, education, and health sciences) into STEM majors (mathematics, physical sciences, biological/life sciences, computer/information sciences, engineering/engineering technologies, and science technologies). NCES further separates these majors into 23 fields of study. The data is shown in Table 1.

Table 1. Percentage of students that switched from non-STEM majors in their first year of study to STEM majors in their last (sixth) year of study (NCES, 2012).

Major field first entered	Percentage of students who transitioned to STEM from field first entered, 4-year programs	Percentage of students who transitioned to STEM from field first entered, CC programs
Social/behavioral sciences	10.9	5.4
Humanities	8.2	9.0
Business	14.1	11.8
Education	10.5	6.2
Health sciences	27.5	8.9

These percentages indicate significant successful transitions and effective pathways from non-STEM to STEM majors. These re-entering students are not part of the pipeline model, as they at one point dropped out of the pipeline. STEP into STEM will explain why this happens, and will recommend policy considerations that will increase the likelihood of these transitions, particularly for URMs/women. As a real example: NCA&T has about 1000 students each year that enter the university undecided.

1.2 BPR Framework/Theoretical Models

Successful transitions and effective pathways into STEM requires a significant investment from HS and postsecondary-education institutions. Institutions must develop a collective-impact, broadening-participation framework, and engage a network of stakeholders that include students, educators, families, community, government, and industry. STEP into STEM will gather data-driven insights to the factors that influence transitions to STEM for undecided/non-STEM students, and demonstrate that certain systematic/programmatic STEM interventions are desired and effective. Characteristics, mechanisms for positive change, and institutional best-practices will be highlighted in support of the framework. Based on this research, recommendations to Higher Education Institutions (HEIs) for improvements and metrics to encourage more URMs into STEM fields will be made.

Theoretical models support this broadening-participation framework, including: Social Cognitive Career Theory, Student Persistence Theory, Human Capital Theory, and Vocational Anticipatory Socialization Theory. The theoretical underpinnings of STEM *pipelines* versus alternative *pathways* is poorly understood (Montmarquette, 2002). This research will also use other studies that have evaluated how students choose their majors (Malgwi, 2005; and Maple, 1991)

The driving factors (input variables) in these theoretical models are largely captured in the survey questions and resulting datasets described in detail later in the Work Plan. The variables in HSLS:09 (High School Longitudinal Study, 2009) (NCES, 2009), and both the BPS:04/09 and BPS:12/17 (NCES, 2016) (Beginning Postsecondary Students Longitudinal Studies, 2004-09 and 2012-17, respectively) are directly related to the theoretical constructs. This research will be guided by these theories, and will use the variables in the longitudinal studies to understand the impacts of key driving forces (determinants of postsecondary outcomes) on the STEM-major decisions of students. Each theoretical model mentioned above is covered in greater detail below, as it pertains to the STEM-major decision process for postsecondary students, particularly those who enter college either undecided or in non-STEM majors.

1.2.1 SCCT: Social Cognitive Career Theory

This research incorporates elements of SCCT that provides a framework for career development and educational decision making. SCCT suggests that career development is influenced over time by supports and barriers which can be both individual and contextual in nature (Bandura, 2001; Lent, Brown, & Hackett, 2000; and Wang, 2013). An individual's combined characteristics like gender, ethnicity, environmental surroundings, and socioeconomic status shape career development and career goals. SCCT theory suggests students develop interest in a subject (i.e. STEM), define goals (i.e. getting a STEM degree or getting educated for a STEM job), and gain confidence (*or self-efficacy*) that the goals are achievable. The SCCT framework implies that students will develop expectations and self-efficacy in math and science through academic experience and performance in HS. Then in postsecondary education, their unique institutional context (choice of HEI) along with academic experiences and performance, as well as economic factors (ability to pay for school) will impact career goal achievement.

1.2.2 Student Persistence Theory and Human Capital Theory

College student persistence models (Tinto, 1998; and Metz, G.W., 2002) inform college STEM decisions. Persistence is also tied closely to human capital theory that suggests decision-making is influenced by a students' perception of productivity-enhancement and investment return (Becker, 1993; and Paulsen, 2002). Many studies have proven strong relationships between the human-capital predictor, academic preparation, and the choice to enter a STEM field. Course-taking and performance patterns directly impact the choice of a STEM major, and the success of earning a STEM degree (Levine, 1991; and Song, 2004).

1.2.3 VAS: Vocational Anticipatory Socialization Theory

VAS theory is about the process of learning about and selecting a career. VAS has been used to study students' perceptions of the most influential sources and content of encouraging/discouraging career messages. VAS suggests that who delivers career messages and what message is conveyed will help boost self-efficacy and outcome expectations, thereby helping students envision themselves in STEM career settings. Results suggest that mothers, followed by teachers/professors, friends, and fathers are perceived as the most influential encouraging VAS sources (Powers, 2016; and Jahn, 2014a).

The types of messages students receive, message sources, and students' career frameworks (enjoyment-based, ability-based, and goal-based) affect their educational and vocational interests (Jahn, 2014b). There are generally two types of VAS messages: *personal fulfillment* (advising students to prioritize their well-being); and *career detail* (advising students about specific aspects of an occupation).

1.3 Barriers, STEM Interventions, and Best-Practice HEIs

1.3.1 Barriers

Barriers to success in STEM are many. SCCT (2000) recognizes that these barriers can serve as blockages in the STEM pipeline, or as barriers into *alternative pathways to STEM.* It is the *alternative pathways* to STEM that this research is studying (some barriers are listed below in Table 2).

Table 2. Barriers to successfully earning a STEM degree

Academic		
Preparation for science education	Irrelevant science curriculum in HS & college	Curricula in HS & college doesn't foster interest
Unclear pathways to STEM success	Too many choices	Hard to understand program/field requirements
Poor advising	Lack of academic support systems	Lack of academic/career goal setting
Chilly climate in research & science at HEI		
Socioeconomic		
Gender discrimination	Positive childhood experiences with science/math	Lack of URM/female role models
Gender role expectations	Masculine view of men in science fields	Opportunity costs transitioning to STEM high
Opportunity costs transitioning out of STEM low (Xie, 2003)	Effects of peers	Career outlook (opportunities, income expected, work conditions, location of job)
Demographic		
Underrepresented minority	Male vs. Female (perceived biological differences) (Blickenstaff, 2005)	Family formation (spouse, children, other)

1.3.2 STEM Interventions

In recent years, high schools, CCs, and 4-year institutions have been exploring systems and programs that might be effective in addressing the STEM barriers detailed above. Systematic implementations in postsecondary institutions include, in part:

- Improved financial-support and institutional-funding systems (crucial for timely graduation) (Kokkelenberg, 2006)
- Enhanced methods for colleges to deliver instruction (academic-program structure)
- Improved student-intake processes
- Implemented student-support-service systems
- Monitored student-progress mechanisms
- Changed pathways from fragmented course-taking to "clear, coherent academic/career pathways" (Completion by Design)

Some states have implemented "Student Success Centers" in some postsecondary institutions. These centers focus on student completion, progress monitoring, support, design of course programs for students, and intake processes. Additionally, starting in 2013, Complete College America (Complete College America, 2014) has supported a systematic program called Guided Pathways to Success (GPS) (White, 2005). The notion behind GPS is that college students are more likely to complete a degree in a timely fashion if they choose a program and develop an academic plan early, have a clear roadmap of the courses they need to complete for a degree, and receive guidance and support to help them stay on plan (Bailey, 2015). GPS has also been employed specifically at a few universities discussed below to improve STEM degree outcomes and metrics. In the Work Plan (section 3.3), a meta-study is discussed that looks specifically at the impacts of GPS for URM/women, and undecided/non-STEM students.

The systematic implementations listed above are usually supported by individual programs. This approach borrows from collective-impact theory, in that it takes the mutually reinforcing activities in many of these programs to result in a systematic approach to improved STEM outcomes. A partial list of "best-practice" programs that are being tested/implemented across the US today include:

- Bridge Programs for HS to CC to 4-year HEIs (including 2+2 approaches)
- Intake Programs:
  - encouraging AP coursework in HS (Moskowitz, 2001)
  - alignment of HEI with HS feeder schools (strategic dual enrollment)
  - use of software to predict course performance based on academic record (useful for guiding students into majors where they have increased success probabilities)
  - template program planning (mapping out degrees, default course registration for 2- or 4-year degree attainment, planned and seamless on-ramps as needed, and milestone requirement definitions to track degree attainment)
  - meta-major programs for undecided students (a large population across the country). Meta-major programs expose undecided students to many majors, while keeping them abreast of STEM fields, for easy STEM transitions during/after the first year.
- Postsecondary Education Programs:
  - Encouraging and supporting enrichment/remedial education in core math and science, if needed (particularly in rural HSs that don't offer advanced math/science to students)
  - Identifying research-experience opportunities
  - Advising (to select major in year 1, advising around course schedules/maps, early-warning advising for milestones missed, automatic-withdrawal advising, and eAdvising)
  - Specifying student-learning outcomes for fields of study
  - Career planning with defined start and end points
  - Scheduling support (aligned to major program mapping, and planned vs. actual results)
  - Monitoring of student outcomes (accelerating completion, preventing wasted credits)
  - Tutoring to support student success

1.3.3 Best-practice HEIs

A review of the literature identified some CCs and 4-year universities that have specific programs devoted to success in STEM. Some of these programs are also focused on URMs/women. The list of HEIs with STEM best-practice programs to support students in STEM include, in part:

- City Colleges of Chicago
- Miami-Dade College
- Florida State University
- Arizona State University
- Valencia College (Orlando, FL)
- New Jersey Council of County Colleges (NJCCC)
- Michigan (independent CCs) – Michigan Student Success Center (MSSC)
- Indiana, Indiana Commission for Higher Education (ICHE)
- New York (CUNY) – Accelerated Study in Associate Programs (ASAP) for CCs
  - Guttman CC- part of CUNY
- California, Long Beach City College

This list is a component of this research, and will be appended as more best-practice HEIs are identified. In the Work Plan, details of a meta-study are discussed. The goal will be to identify top-, middle-, and lower-tier HEIs as it pertains to STEM-program interventions, and to use the NCES datasets to study impact and effectiveness. The list of institutions above will likely fall into the top-tier, best-intervention HEIs. Since the HSLS:09 and BPS:12/17 datasets includes data from the recent decade, the data contained within those datasets will be particularly informative to the effectiveness of recent STEM programs at the universities listed above; all of which began significant STEM interventions since 2012-13. More details on the meta-study are given in the Work Plan below.

1.4 STEP into STEM Theory of Change (TOC) Model

The following TOC Model builds upon socioeconomic models discussed above, and prior research. STEP into STEM seeks to understand *why change happens*, as shown in the lower right of Figure 1.

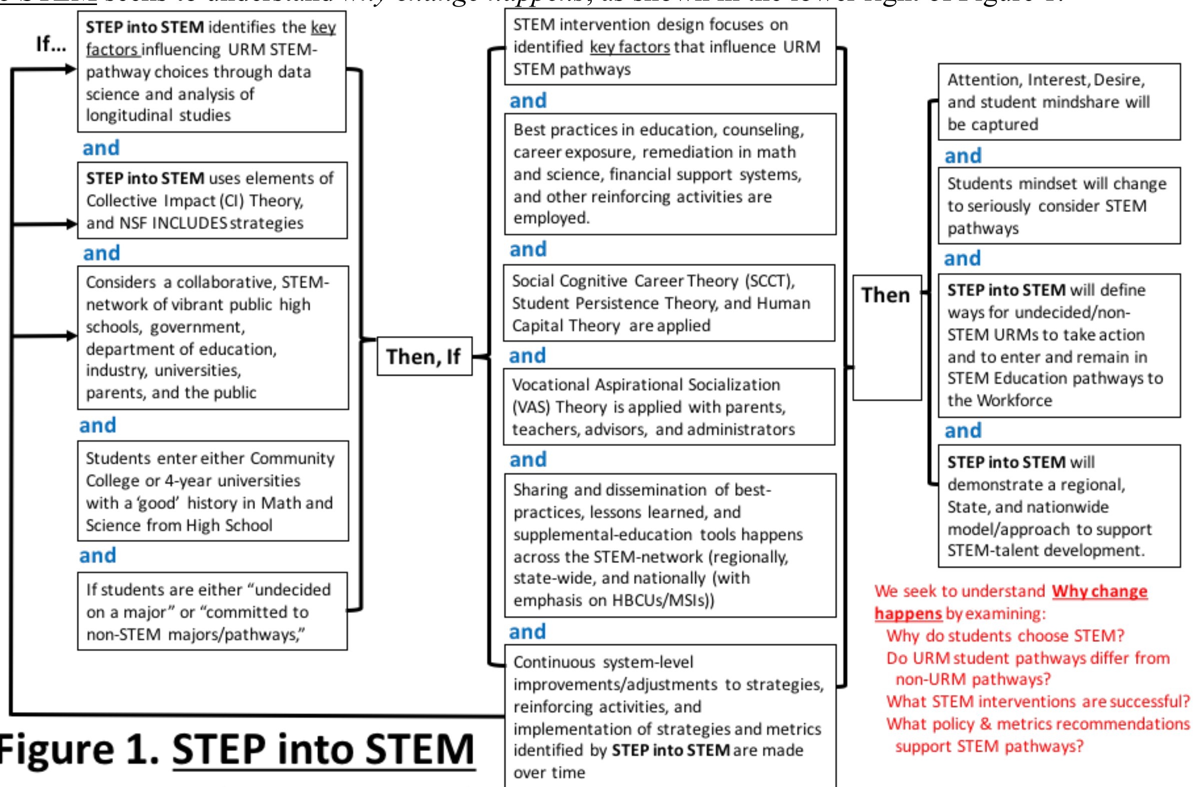

Figure 1. STEP into STEM Theory of Change Model

1.5 Hypotheses Questions

STEP into STEM will address the following questions:

- What are the academic, socioeconomic, and demographic supports and barriers for URM, undecided/non-STEM major transitions to STEM?
- What effective pathways and HEI systems/programs support URM, undecided/non-STEM successful transitions to STEM?
- Which HEIs have data/evidence to support their best-practice, intervention systems/programs?
- What implementable strategies will support successful transitions and effective pathways into STEM, particularly for URMs and undecided/non-STEM students from HBCUs/MSIs?
- What shared measurements/metrics are recommended for HEIs to use nationally that would be useful in the assessment of successful transitions and effective pathways for URMs into STEM?

1.6 Intellectual Merit statement

This research encompasses the potential to advance knowledge in understanding, designing, implementing, and assessing successful transitions and effective pathways into STEM education, particularly for URMs/women who have entered postsecondary education either undecided on their major, or in a non-STEM major. The research will highlight the opportunities to recruit postsecondary students into STEM pathways, when those students did not have firm STEM educational and career commitments upon entry to college. The outcomes of the research will contribute to the state of the art in

STEM education through understanding how STEM best-practice interventions at both community college and 4-year HEIs can broaden participation of URMs in STEM, thereby leading to a national model tailored to identify, motivate, prepare, attract, and retain students in STEM. The research output will provide guidance to policy makers at HEIs across the nation on effective programs to increase URMs in STEM pathways to graduate school and to the workforce. Additionally, data-science methods will be employed to evaluate big datasets from the NCES. Data-science approaches with big datasets will inform researchers-to-follow with state-of-the art methods to find trends, influencing factors/variables, combinations of factors and their impacts, and to define a core set of metrics for STEM transitions and pathways that could be used nationally.

2 Partnerships and Team

A STEP into STEM team primarily from NCA&T combines: the PI, knowledgeable in Collective Impact theory and presently funded by NSF INCLUDES (award #1744477); and three co-PIs representing the College of Education, Social Science, and Data Science. An external evaluation will be performed by SEI (Strategic Evaluations, Inc.) who has performed similar assessments for nationally-funded education grants. Since this research is aimed at BP in STEM, with a focus on transitions from undecided and non-STEM majors, it is important to understand the elements of Collective Impact theory (common vision and agenda, continuous communication, mutually reinforcing activities, shared measurements and metrics, and a backbone organization). Alternative pathways to STEM will succeed if Collective Impact theory is applied. Results of this research will be shared with a network of people, all of whom must be engaged to achieve the vision stated above (section 1.1).

Dr. Greg Monty, PI, NCA&T, is Director of the Center for Energy Research and Technology, focused on energy efficiency and sustainability research. He has extensive academic, industrial, and business experience. He is the PI of the NSF INCLUDES Pilot (EMERGE in STEM) focused on BP of grades 4-12 underrepresented groups in Guilford County into STEM pathways in education to the workforce. The goal of EMERGE in STEM is to give students exposure to STEM careers that will stimulate their interest and entry to STEM pathways. Dr. Monty will lead the project and provide project management expertise (developed throughout industrial career and particularly at Motorola).

Dr. Caroline Booth, co-PI, NCA&T, is an Associate Professor and Interim Chair of the Department of Counseling within the College of Education. She also serves as the Executive Director of the Center for Behavioral Health and Wellness, a funded student success initiative on campus. This Center currently works to ensure academic success of undergraduate students through targeted interventions to increase student engagement. Her main research interests are social science, career development, counseling, wellness over the lifespan, and educational assessment (having served as educational/assessment coordinator on prior NSF-funded research). Dr. Booth will lead social-science efforts to understand the determinants of postsecondary STEM outcomes, based on theoretical models.

Dr. Marwan Bikdash, co-PI, NCA&T, is Professor and Chair, Department of Computational Science and Engineering. He founded and is interim Director of the Visualization and Computation Advancing Research (ViCAR) supercomputing Center. He has been actively engaged in teaching and research activities in data science, complex systems, computational-system theory, signals and systems, and computational intelligence. He has published >130 journal and conference papers, and has managed more than $4M of research funding as Lead PI. His research has been supported by NSF, NASA, several DOD agencies, HP, and others. He is currently conducting research in the vulnerability of infrastructure especially as can be deduced from GIS datasets. Dr. Bikdash will be involved in the overall design, formulation, and study of the research questions to be asked, developing rubrics for institution and program evaluation, conducting exploratory statistics and predictive modeling on the data, especially using regression analysis, principal component analysis, and time-series models.

Dr. Dukka KC, co-PI, NCA&T, is an assistant professor and graduate coordinator, Computational Science and Engineering Department at NC A&T. His research interests are development of algorithms and tools to analyze biological data, and data science. He uses various machine-learning and data-mining tools to mine important information from biological data. He has been instrumental in developing big-data infrastructure at NCA&T by serving as the first ever XSEDE Campus Champion from the University; he will be involved in the data-science aspects of this project.

Strategic Evaluations, Inc. (SEI, under the direction of Dawayne Whittington) will conduct our external evaluation (section 3.3.1)**.** This minority-owned business is committed to increasing diversity of scientists. The firm brings 15 years of project evaluation experience that improved retention of URMs in science. SEI has led the evaluation of more than 70 different grants designed to increase diversity in the sciences, including several NSF-funded HBCU-UP, TIP, and Noyce programs. SEIs data collection will be vital for formative and summative assessment as well as helping the team meet the research aims.

In addition to the core team above, support from NCES has been, and will continue to be useful to understand the datasets from HSLS and BPS. Elise Christopher, elise.christopher@ed.gov, is the Project Officer for HSLS. David Richards, david.richards@ed.gov, is the Project Officer for BPS. Detailed communications have been ongoing with these partners.

3 Objectives, Metrics, Logic Model, and Work Plan

3.1 Specific Objectives

Based upon the goals stated above (section 1.1.1), the specific Objectives of this research include:

- **Objective 1: Analyze data from longitudinal studies** (from NCES) to understand the factors that influence STEM transitions and pathways, particularly for those URM students that enter postsecondary education either undecided or in non-STEM majors.
- **Objective 2: Study the effectiveness of STEM interventions** (using NCES datasets) by evaluating HEIs that have significant efforts/investments in systems/programs intended to support STEM pathways, as compared to HEIs that have not made such programmatic changes.
- **Objective 3: Produce recommended strategies** (including measurements/metrics) based on evidence and knowledge that supports postsecondary HEIs, particularly for URMs, to successfully and effectively broaden participation in STEM pathways by drawing upon and recruiting from the pool of undecided or non-STEM majors entering their institutions.

3.2 Logic Model for STEP into STEM

In support of the research Goals (section 1.1.1), Objectives (section 3.1), and aligned with the TOC Model (section 1.4), a Logic Model for STEP into STEM (Table 3) was developed. Immediate outputs, intermediate outcomes, and longer-term outcomes are detailed. STEP into STEM is primarily impacting immediate outputs, as intermediate and longer-term outcomes will be expected results after the research has understood the determinants of postsecondary outcomes and made recommendations on policy/metrics. A major section of the Logic Model on the Work Plan is further explained in Table 4.

Table 3: Logic Model for STEP into STEM

Inputs	Activities/ Work Plan	Outputs	Mediating Individual/Critical Factors	Intermediate Outcomes	Mediating Organizational Factors	Longer-term Outcomes
Promising Students		Pathways for undecided and non-STEM students toward STEM are characterized	Student self-efficacy and confidence lacking	Enhanced academic performance of Students in STEM	Time required for High Schools and HEIs to make changes in their systems/curricula/programs	Capabilities and Capacity of Students in STEM is increased
Varied Educational Experience		Key Factors that represent barriers to success are detailed	STEM knowledge and skills matched to actual jobs	Students select alternative pathways to STEM to explore	Budgetary constraints on public education institutions	Increased Number and Percentage of URMs in STEM Pathways
Prior Career Guidance		Influences specific to URM STEM success are specified	Financial support lacking	Students that begin on STEM pathways continue on STEM Pathways	Successful STEM interventions may be institution dependent, and will not be "one size fits all."	Increased Number and percentage of URMs enter STEM workforce, and remain in Workforce
Perceptions of Options for their pathways		HEIs are characterized by a rubric as to their level of commitment to system/program changes for STEM success	Parental/peer support lacking	Positive attitudes toward STEM		Regional/State/National community use best practice model from STEP into STEM
Facilities to support STEM activities	See STEP into STEM Work Plan in Table 4.	Data Analtics are produced that establish links between STEM interventions and effective pathways to STEM	Family structure inhibits success	HEIs establish substantive systems and programs that encourage students to STEM pathways		STEP into STEM model implemented for URMs across the HBCU/MSI community nationally
Identified STEM Workforce needs		A comprehensive set of recommendation emerge that could propel HEIs to implement successful STEM systems/programs nationally.		Increased support from across community		Local/Regional/State/National community and education systems increases capacity/capability
STEM-concerned community members				Capacity building begins to take shape across community (industry interaction, state-wide education organizations, others)		
Socioeconomic status of many URMs is challenged				High School feeder systems identify systems/programs that support alternative pathways to STEM		
				Parents/Teachers/Administrators learn what/how to communicate STEM pathway opportunities to students		
				Increased attention, interest, desire, and action by Students in STEM pathways		

3.3 Work Plan

The STEP into STEM Work Plan in Table 4 details specific objectives/tasks/steps along the timeline from pre-award through 3 years of research.

Table 4: STEP into STEM Work Plan (Tasks in support of Project Goals and Objectives)

Objective (Leaders)	Task	Pre-award	2018	2019			2020			2021	
			Fall	Spring	Summer	Fall	Spring	Summer	Fall	Spring	Summer
1: Analyze data from longitudinal studies (Dr. Marwan Bikdash, Dr. Dukka KC, Co-PIs)	1: Acquire Longitudinal Data	Apply for Restricted Use Data (RUD) License from NCES (reference added)									
			Import HSLS:09 RUD	Understand dataset							
			Import BPS:04/09 RUD	Understand dataset							
			Import BPS:12/14 RUD	Understand dataset							
				Import BPS:12/17 Public Use Data	Understand dataset						
						Import BPS:12/17 RUD	Understand dataset				
	2: Analyze Longitudinal Data		Develop algorithms with HSLS and BPS:04/09 datasets				Improve Data Analytics Methodologies				
					Data Science studies of HSLS:09 and BPS:04/09 (Trends, Influencing factors, splits for HBCU/MSI vs other HEIs			Repeat Data Studies with Improved Data Analytics			
						Repeat Data Studies on BPS:12/17 datasets			Repeat Data Studies with Improved Data Analytics		
2: Study the effectiveness of STEM interventions (Dr. Caroline Booth, Co-PI)	1: Perform Meta-study		Step 1: Develop rubric to classify HEIs for STEM intervention systems/programs		Assign HEIs to Tiers based on rubric						
					Step 2: Develop analytics to evaluate HEIs based on Tier and STEM pathway success			Repeat Data Studies with Improved Data Analytics			
3: Produce recommended strategies/metrics (Dr. Greg Monty, PI)	1: Recommendations and metrics									Generate Report 1 (see proposal)	
										Generate Report 2 (see proposal)	
										Generate Report 3 (see proposal)	
										Generate Report 4 (see proposal)	
										Generate Report 5 (see proposal)	
External Evaluation (SEI)					Interim Evaluation Report			Interim Evaluation Report			Final Evaluation Report

Tasks in support of the Objectives are:

<u>Objective 1, Task 1</u>: Acquire longitudinal data.

Innovative analytical techniques including Random Forest, Principal Component Analysis, XGBoost, and Deep Learning will be explored to analyze the large datasets from NCES. NCA&T hosts the Visualization and Computation Advancing Research Center (ViCAR), with super-computer computation, extensive data storage/backup, and visualization capabilities (see the Facilities, Equipment, and other Resources attachment to this proposal for details).

Data will be acquired primarily from a few longitudinal studies completed by the NCES. One study used will be HSLS:09 (High School Longitudinal Study of 2009). This study collected data from 23,000+ 9th graders from 944 schools in 2009, conducted follow-up data collection in 2012 (as seniors), and again in 2016 (4 years after HS graduation). The study includes secondary/postsecondary years. The focus of HSLS:09 was students' trajectories from beginning HS to postsecondary education, workforce, and beyond. HSLS:09 tracked the majors/careers students pursued (when, why, and how). It also asked a critical question of "how do students choose STEM courses, majors, and careers?" For HSLS:09, all data is available today, except the second follow-up in 2016, which is expected to be released in early 2018 (before STEP into STEM will begin). Variables used in HSLS have been defined (NCES, 2009).

Data will also be acquired from the NCES Beginning Postsecondary Students Longitudinal Studies; BPS:04/09 and BPS:12/17. BPS:04/09 data was collected from 16,700 nationally-representative postsecondary students/institutions. BPS datasets are a combination of data from: NPSAS (National Postsecondary Student Aid Study, FAFSA, ACT, IPEDS (Integrated Postsecondary Education Data System), and NSLDS (National Student Loan Data System). Initial data was captured at the end of the students' first academic year, 2003-04, then at the end of the students' third (2005-06) and sixth (2008-09) year. There are many variables (~650) used in BPS as documented in (NCES, 2016). BPS:12/17 has completed data collection, and David Richards from NCES has stated that the public data will be released in early 2019, with full release of the restricted-use data (RUD) file in the fall of 2019. The BPS:12/14 RUD file was released in May 2016. Therefore, this project will have access to sufficient preliminary data, and the final follow-up BPS:12/17 data will be available about 1 year and 3 months after STEP into STEM begins. With the BPS:04/09 and BPS:12/14 RUD files, any data-analysis algorithms and tools developed can be exercised, and will be ready for final BPS:12/17 RUD file release.

BPS collects data on student persistence in, and completion of, postsecondary education programs, their transition to employment, their demographic characteristics, and changes over time in their goals, marital status, income, debt, among other indicators.

The rich temporal BPS datasets contain critical information for this research effort, including in part: choices/timing/changes of majors; influence of HS grades/courses/test scores; postsecondary coursework taken (both STEM and non-STEM); GPAs in postsecondary education (both STEM and non-STEM fields); parental influence; ethnicity/sex; specific HEIs students attended; selectivity of institution attended; financial aid influences, and degree programs completed.

Importantly, the HSLS and BPS datasets will allow comparison of performance/trends at HBCUs/MSIs, to other HEIs, and comparisons between URMs and non-URMs.

Pre-award activity will include acquiring a RUD License from NCES (NCES, 2017). To receive a RUD License, applicants must designate a Principal Project Officer (PPO), Senior Official (SO), and Systems Security Officer (SSO) (must sign the required documents for the License). The PI will insure that all documents are provided to NCES to gain access to the HSLS and BPS longitudinal restricted-use data, before the grant begins. Datasets will be supplied from NCES on DVDs in flat-file, csv formats.

Objective 1, Task 2: Analyze longitudinal data.
The key influencers on student choice of major, and success on STEM pathways are academics, socioeconomics, and demographics. STEP into STEM will look at which aspects of these three influencers impact student pathways. The BPS datasets will provide specific, student-by-student data on when they chose a major, if/when they transitioned to a different major, and which major they ended in after 6 years. This will be studied at CCs and 4-year HEIs, with additional data from HSLS that includes HS factors into postsecondary years and beyond.

Examples of variables available in the BPS datasets are shown in Table 5. These variables, and many more, will give insights into driving forces that effect transitions students make from undecided/non-STEM to STEM.

Table 5. Some NCES BPS variables of interest for STEP into STEM by significant influencer category

Academic		
ACT/SAT scores	Level/control of HEI first attended	HS GPA
Degree program selected Y1	Selectivity of HEI first attended	Highest math level HS, Y1, Y6
Degree program Y6	Cumulative persistence in STEM	High School degree type
GPA STEM/non-STEM Y1, thru Y6	Remedial courses taken	Entered STEM but left by Y6, attainment of degree
STEM credits attempted Y1, thru Y6	% credits earned Y1 that were STEM, thru Y6	Year of entrance into specific field of study
STEM credits earned Y1, thru Y6	% withdrawn/failed STEM attempted Y1, thru Y6	Total credits earned in specific fields of study thru Y6
Use of institutional services: academic advising/support; career; financial aid; health services		
Socioeconomic		
Pell Grant recipient	Income level Y1 thru Y6	support from peers/parents
1st family member in college	Marital status	siblings who attended college
Demographic		
Highest degree of parents	Race/ethnicity	Sex

The variables included in the NCES studies are supported by the success factors covered in research literature. They include: AP coursework, math ability, gender, ethnicity, HS GPA, college experiences (Kokkelenberg, 2010), early childhood interest in STEM (middle school and leading up to postsecondary education) (Xie, 2003), continuing or entering a STEM major in either year one or two of postsecondary education (Xie, 2003), long-term interest in STEM (Kilgore, 2007), and for engineering majors: an interest in engineering, science, and math. These variables are inputs to the BPR Framework/Theoretical Models (section 1.2).

Additionally, typical recommended measures of individual success for STEM degree achievement include: postsecondary GPA, graduation rates, persistence, completion time or time to degree, certification in some sub-field, employment, and post-graduation earnings. Most of these measures are covered in the NCES datasets and will inform this research.

Objective 2, Task 1: Perform meta-study.

The meta-study will look at effectiveness of STEM interventions at CCs, and 4-year HEIs. This task will examine the strengths and weaknesses of institutional systems/programs established to encourage more students to enter STEM pathways through graduation and to the workforce.

The barriers that students face as they consider STEM were described above in section 1.3.1. Additionally, the types of STEM interventions that have been employed by HEIs across the nation were described in section 1.3.2. Finally, a non-exhaustive list of universities that are considered, based on the literature, to have incorporated significant STEM intervention programs were listed in section 1.3.3. Since the HSLS and two separate BPS datasets (BPS:04/09 and BPS:12/17) have information on individual HEIs across the country, a meta-study will evaluate the effectiveness of various STEM systems/programs.

The approach to be taken will include the following steps:

Step 1: Classify HEIs into Tiers, based on best practices at the institute level. A developed rubric will classify HEIs based on types of programs/systems/best-practices. At least 3 Tiers will be established (Tier 1: significant STEM intervention systems/programs, to Tier 3: few STEM interventions occur).

Step 2: Once the universities have been classified into Tiers, a study will proceed to determine whether students at those Tier levels truly benefitted from the interventions in place. Students will be identified by the Tiers and HEIs attended as noted in the NCES datasets. NCES variables will be used to determine which students benefited most from the interventions. A specific focus will be placed on URM transitions from undecided/non-STEM majors into STEM majors.

Data-Science and Data Analytics Methodologies to be applied to Objectives 1 and 2:

Innovative Data Science Questions

1. Identify the most important variables that can predict whether students with a given profile will "step into STEM" in one sense or another. This is a classification approach: Compare the results with factors identified using importance analysis based on a regression approach.
2. Developed predictive models of reduced dimensions by combining factors linearly or nonlinearly
3. Test whether the factors identified in the literature as major predictors are indeed so, in the sense that predictions based on them are not significantly inferior to predictions based on the full data.
4. Test whether temporal data is rich enough to support time series models to predict future trends.

Detailed discussion of the classification problem

As an example of our approach, state-of-the-art machine-learning algorithms like Random Forest (RF) will be used to predict whether a student will step into STEM and will do so through a model of reduced dimensionality. RF is equipped with a mechanism that can deal with the high dimensionality of data by selecting only the important features. This is related to Collinearity analysis.

The variables used in the NCES study can be viewed as features and whether the student graduates from the STEM field or not as the outcome/label. Given the vast amount of data, about 650 variables (features), and the outcome (whether they graduated from STEM field or not), this problem can be formulated as a binary classification problem. We will also use Random Forest (an ensemble based approach) to develop a predictive model.

Random Forest is a classification scheme, originally proposed by (Breiman, 2001), constituted by an ensemble of decision trees that are grown using only some subsets of features. The final classification is obtained by averaging the decisions of all trees in the forest. RF has been applied in various scientific disciplines. Co-PI KC has successfully applied RF to various bioinformatics problems (Ismail, 2016a; Ismail 2016b; and Ismail 2015).

Based on this experience, RF will be a tool to identify important variables pertaining to STEM degree attainment (Belsley, Kuh, and Welesh, 1980; Bishop, Fienberg, and Holland, 1975; and Breiman, Friedman, Olshen, and Stone, 1984).

RF is known for its lower generalization error, relative robustness to outliers and noise, and immunity against over-fitting compared to other machine-learning methods (Jo, 2014; Pedregosa, 2011). The randomness of bootstrap sampling enhances prediction accuracy. Each individual decision tree in the random forest is constructed with a bootstrap sample from a training dataset. The sample is made up of a root node, internal nodes, and leaves. Each node represents a feature that is selected based on a criterion. A node may have two or more branches. Each branch corresponds to a range of values for that selected feature. The node branching of the decision tree is performed by computing the Gini impurity index for each feature and only the most important feature that splits the training data into the purest classes is selected. A node and its ranges of values, that split the sequences, are chosen as decision rules.

Variable (Feature) importance and feature selection:

In RF, the Gini impurity index for each feature is calculated and considered for node split. The importance of the features is estimated as the sum of the Gini index reduction (from parent to child) over all nodes in which the specific feature was split. The feature that contributes to the largest Gini index reduction is the most important. Thus, the features can be ranked according to their importance. Using the Gini index, initially, variables (features) will be ranked based on the importance for STEM degree attainment. Once this step is performed, the top-n (n: 50, 100) variables (features) will be used to study other relevant research questions.

In the context of regression analysis, a similar concept consisting of finding random regression trees can be applied. The above approach will be complemented by additional exploratory statistical analysis such as the Principal Component Analysis (PCA) (Jolliffe, 1986), and XGBoost (Chen, 2016). PCA is based on finding patterns in the factor covariance matrices, and is particularly important to identify "principal components" (linear combinations of the explanatory variables) that can have high explanation power. Nonlinear principal component analysis (Kramer, 1991) can be used to identify and remove correlations among problem variables as an aid to dimensionality reduction, visualization, and exploratory data analysis.

Model Validation:

Model validation will assess the models thoroughly for prediction accuracy. In this study two evaluation strategies will be adopted; ten-fold cross validation and independent test samples.

Ten-fold Cross-validation:

Ten-fold cross validation is a model validation technique to assess how the results of a model will be generalized to an independent data set. In ten-fold cross-validation, the data set is divided into ten folds (parts): nine folds are used for training, and a single fold is used for testing.

Independent test sample:

An independent test sample is a set of data that is independent of the data used in training the model. In addition to 10-fold cross-validation, part of the dataset (20%) is set aside for independent test samples which will be used to evaluate the classification model as well. Once the predictive model is developed, it can be used to predict whether a student is likely to graduate from a STEM field given the variables.

Objective 3, Task 1: Produce Recommended Strategies

Based on extensive analysis of HSLS and BPS datasets, and a meta-study of effectiveness of STEM intervention systems/programs implemented at universities across the nation, evidence will be produced to guide policy makers at HEIs. Results of the research will be organized into individual reports related to the following topics (with subsections broken out to specifically address the different factors related to URMs vs. non-URMs; and HBCUs/MSIs vs. other HEIs):

- **Report 1:** Academic, socioeconomic, and demographic barriers for undecided/non-STEM major transitions to STEM;
- **Report 2:** Effective pathways and HEI systems/programs that support undecided/non-STEM successful transitions to STEM;

Report 3: HEIs of Best Practice supporting undecided/non-STEM transitions to STEM (a report on Tier classification);
Report 4: Recommended implementable strategies to support successful transitions and effective pathways into STEM, particularly for URMs and undecided/non-STEM students. (a focused report for HBCUs/MSIs); and
Report 5: Recommended metrics for HEIs to share on a national basis to assess the successful transitions and effective pathways into STEM.

(This last report may leverage research literature ideas for metrics that include, in part: 1. the number and percentage of successful completed degrees, by total and by demographics; 2. the number of STEM programs offered with high employment opportunities; 3. credits completed that apply to or transfer into a major or program; 4. the number of students graduated and employed in a field of study; 5. financial aid awarded to students (including loans) referenced to earnings potential in the field of study; and 6. others.

3.3.1 Evaluation, Assessment and Progress Indicators

The evaluation team's approach will center around two key strategies. The team will first design an ongoing TextIt campaign (https://textit.in/) for electronically collecting progress data every 4-6 weeks from core team members contributing to the research project. Secondly, the team plans to conduct annual interviews with project leadership to complement to data collected via the ongoing TextIt campaign. The two strategies will combine to serve both the formative and summative assessment needs of project.

Ongoing Text-Message Check-Ins with Core Team Members – The evaluation team will incorporate an ongoing TextIt campaign as a core strategy for monitoring the implementation of the work plan presented in the proposal. TextIt is an Internet-based platform that will enable the evaluation team to send team members associated with the work plan SMS-based prompts that they can quickly respond to from their mobile devices. As participants respond via text, every interaction creates a data point on the TextIt platform that is automatically associated with each team member, allowing for real-time analytics that will be shared in aggregate with the project leaders. For deeper analyses, the evaluation team plans to import data into Excel and SPSS. The TextIt campaign the evaluation team will deploy is structured to automatically send each team member a series of short open- and closed-ended questions designed to document their individual progress on benchmarks outlined in the proposed work plan. The data will be collected anonymously every 4-6 weeks by the evaluation team and will require less than 10 minutes of team members' time to reply via mobile device. The TextIt campaign will be designed to focus on a series of six questions that are critical for independently reporting the research progress by the team.

Sample TextIt Campaign Questions

1. How likely are the objectives (for the current semester) to be achieved according to the timelines specified in the work plan? Is the project ahead, on, or behind schedule?
2. What factors/activities (both technical and non-technical) have been key in keeping the progress on schedule, ahead of schedule, or behind schedule?
3. Should the timeline for completing the research objectives (this semester) be revisited?
4. Do team members have adequate resources (money, equipment, facilities, training, etc.) to achieve the objectives for the semester?
5. Are the long-term objectives still realistic considering progress being made?
6. Should overall priorities be reviewed/changed to more efficiently achieve the objectives?

Annual Interviews with Core Leaders - The evaluation team will interview core members of the research team annually. The protocols will be developed based on the results of the ongoing TextIt campaigns. Interview questions will be designed to help the evaluation team better understand the progress the team is making as well as more easily document recommendations that are shared across the team.

Interim Reporting - The evaluation team will electronically share aggregate descriptive datasets after each collection with project leaders, as well as summarize the collective sets of data for annual interim reports. All interview data and open-ended comments from TextIt campaigns will be transcribed, coded, and analyzed in *Atlas.ti (*a qualitative analysis software program), and themes will be reported in the annual interim reports. The descriptive datasets and annual reports are expected to assist project leaders in: 1) identifying trends regarding the progress (or lack thereof) toward objectives, 2) reviewing recommendations related to the progress and priorities, and 3) determining what, if any, actions need to be taken to maintain progress as included in the proposed timeline. The interim reporting will also look to determine lessons that can be learned from these monitoring efforts to improve future planning and monitoring activities.

Final Evaluation Reporting - At the end of Year 3, the evaluation team will prepare a summative report. The summative evaluation report will not only document the success and challenges the project leaders faced in meeting the stated outcomes, but also look to document the early contributions the leaders' work is making to the field. Additional variables such as documenting the extent to which the five reports are downloaded as well as cited by others in published papers are examples of how the evaluation team will document the early "reach" of the team's work.

4 Potential for Impact and Scale

Impact and scale were designed into STEP into STEM to broaden participation nationally. The reports disseminated from STEP into STEM will give clear policy recommendations that will be implementable at HBCUs/MSIs and other HEIs. A detailed set of shared measurements/metrics that might be a common best-practice for the nation's HEIs will be generated. A common and shared set of metrics is a central element of collective-impact thinking, and STEP into STEM will have the shared metrics as a key goal in the research output. Additionally, there will be recommendations that will apply at the HS level.

The data analytics algorithms and tools will be documented and available to other researchers that wish to use them for other studies that involve HSLS or BPS datasets, and others. The creative methodologies developed will make exploration of big datasets from NCES more feasible. The software developed will be available to any research group, with support from the STEP into STEM team.

4.1 Dissemination of Results

The reports to be generated will be made freely available. Additionally, journal articles and conference papers will be written for publication. Since the policy recommendations/metrics are expected to be impactful to HBCUs/MSIs, STEP into STEM will engage the leadership of those STEM-oriented institutions to discuss how to implement the suggested approaches. A common practice and metrics across HBCUs/MSIs will likely bring many more URM students into STEM pathways to the STEM workforce. Additionally, since NCA&T is the largest STEM/engineering HBCU in the nation, STEP into STEM will engage the Chancellor and staff to discuss how best practices can be implemented locally on campus. By engaging NCA&T leadership in the discussion, the severe underrepresentation of African American students in engineering/geosciences/physics at the UG level can be addressed, and can be openly communicated within the community of HBCUs/MSIs.

5 Broader Impacts of the Proposed Work

STEP into STEM has high potential to broaden participation in STEM because NCA&T will focus on successful transitions and effective pathways for undecided/non-STEM postsecondary education students to join and remain on STEM pathways to the workforce. The societal benefits and outcomes from this STEM Pilot include, in part:

- increased participation of URMs/women in HBCUs/MSIs across the nation *(best practice systems and programs will be identified and disseminated to HBCUs/MSIs that will increase the number/percentage of STEM graduates at those HEIs)*;
- improved STEM education and educator development for secondary and postsecondary education *(educators and institutions will be given recommendations of best practices to implement)*;
- increased public understanding and engagement related to science and technology pathways for all students, but particularly URMs/women *(through dissemination of results, the research intends to help the community and parents, particularly, understand STEM pathways and deliver synchronous messages and experiences to students related to their entry opportunities into STEM careers)*;

- improved career development and well-being of individuals in society *(selection and continuation of STEM in college toward the STEM workforce is the goal of this research, and students will directly benefit)*;
- development of a diverse, globally competitive STEM workforce *(URMs/women are the focus of the research and will directly benefit)*;
- increased partnerships between academia, industry, and other stakeholders *(recommendations to emerge from the research will draw together high school feeder programs, CCs, and 4-yr institutions, and particularly the HBCUs/MSIs across the nation to establish systems/programs and measurements/metrics that will help to broaden participation nationally)*; and
- enhanced infrastructure for research and education *(the new data-science methodologies/algorithms developed will be useful in future research on BP, and particularly with the big-data, longitudinal-study datasets that are rich with information)*.

6 Results from Prior NSF Support

The following existing and prior support from NSF has pertinent alignment to this BPR proposal.

NSF/INCLUDES - 1744477 (1/2018-12/2019, $300k) "NSF INCLUDES DDLP: EMERGE in STEM (Education for Minorities to Effectively Raise Graduation and Employment in STEM)" (PI: Greg Monty). Intellectual Merit: Innovative and transformative collective-impact model for STEM education, grades 4-12. Focused on how career knowledge and exposure can broaden participation of underrepresented minorities (URMs) in the STEM pipeline/pathways to the workforce. Broader Impacts: EMERGE in STEM engages URMs/women (~15,000+ students). Societal benefits/outcomes will include: educating Deaf students; helping grade 4-12 teachers and administrators incorporate additional STEM tools; bringing parents/community directly into education process; and increasing collaboration between public, private, and university organizations.

NSF/EEC - 1242139 (9/2012-8/2014, $199k) "NUE: Enhancing Undergraduate Students' Learning and Research Experiences through Hands on Experiments on Bio-Nanoengineering" (PI: Narayan Bhattarai, Assessment/Education Coordinator: Caroline Booth). Intellectual Merit: This project focused on increasing knowledge and research related to bioengineering through the development of engineering education best practices. Broader Impacts: Project resulted in the development of an easily replicated educational sequence to enhance undergraduate student learning and research experiences in bio-engineering and bio-nano devices and systems. The societal outcome was an expansion of the number of students exposed to and interested in nanotechnology.

NSF EAGER - 1647884 (8/2016-8/2018, $149.9k) "EAGER: A novel approach to improve template-based multi-domain protein structure prediction" (PI: KC). Intellectual merit: The innovative proposal aims to improve performance of template-based modeling approach for multi-domain protein structure prediction by: a) applying novel threading algorithms and b) using innovative ideas to identify distant or non-homologous template structures. Broader Impacts: This project will develop research capabilities in bioinformatics at NCA&T, and bolster research opportunities to minority students.

NSF ABI Development – 1564606 (06/2016-05/2020, $144.6K) "Integrated platforms for protein structure and function predictions" (PI: KC) Intellectual merit: The project will enhance the performance of protein structure prediction pipeline by using innovative GPU based parallelization. Broader Impacts: This project will develop a widely-used protein structure prediction pipeline, and will train undergraduate and graduate students (including underrepresented minorities and women from NCA&T, the nation's largest HBCU). New courses on bioinformatics and computational biology will be developed at the NCA&T Computational Science and Engineering Department.