2018 ASEE Annual Conference & Exposition
June 24-27, 2018 Salt Palace Convention Center Salt Lake City, UT
Paper ID #21572

Fundamental: Optimizing a Teacher Professional Development Program for Teaching STEM with Robotics Through Design-based Research

Dr. S.M. Mizanoor Rahman, New York University

Mizanoor Rahman received Ph.D. degree in Mechanical Engineering from Mie University at Tsu, Japan. He then worked as a research fellow at the National University of Singapore (NUS), a researcher at Vrije University of Brussels (Belgium) and a postdoctoral associate at Clemson University, USA. He is currently working as a postdoctoral associate at the Mechanical and Aerospace Engineering Department, Tandon School of Engineering, New York University (NYU), NY, USA. His research and teaching interests include robotics, mechatronics, control systems, electro-mechanical design, human factors/ergonomics, engineering psychology, virtual reality, artificial intelligence, computer vision, biomimetics and biomechanics with applications to industrial manipulation and manufacturing, healthcare and rehabilitation, social services, unmanned autonomous vehicle (aerial and ground, indoor and outdoor) systems and STEM education.

Mrs. Veena Jayasree Krishnan, New York University

Veena Jayasree Krishnan received a Master of Technology (M. Tech.) degree in Mechatronics from Vellore Institute of Technology, Vellore, India in 2012. She has two years of research experience at the Indian Institute of Science, Bangalore, India. She is currently pursuing Ph.D. in Mechanical Engineering at NYU Tandon School of Engineering. She is serving as a research assistant under an NSF-funded DR K-12 research project to promote integration of robotics in middle school science and math education. For her doctoral research, she conducts mechatronics and robotics research in the Mechatronics, Controls, and Robotics Laboratory at NYU.

Dr. Vikram Kapila, New York University

Vikram Kapila is a Professor of Mechanical Engineering at NYU Tandon School of Engineering (NYU Tandon), where he directs a Mechatronics, Controls, and Robotics Laboratory, a Research Experience for Teachers Site in Mechatronics and Entrepreneurship, a DR K-12 research project, and an ITEST research project, all funded by NSF. He has held visiting positions with the Air Force Research Laboratories in Dayton, OH. His research interests include K-12 STEM education, mechatronics, robotics, and control system technology. Under a Research Experience for Teachers Site, a DR K-12 project, and GK-12 Fellows programs, funded by NSF, and the Central Brooklyn STEM Initiative (CBSI), funded by six philanthropic foundations, he has conducted significant K-12 education, training, mentoring, and outreach activities to integrate engineering concepts in science classrooms and labs of dozens of New York City public schools. He received NYU Tandon's 2002, 2008, 2011, and 2014 Jacobs Excellence in Education Award, 2002 Jacobs Innovation Grant, 2003 Distinguished Teacher Award, and 2012 Inaugural Distinguished Award for Excellence in the category Inspiration through Leadership. Moreover, he is a recipient of 2014-2015 University Distinguished Teaching Award at NYU. His scholarly activities have included 3 edited books, 9 chapters in edited books, 1 book review, 61 journal articles, and 140 conference papers. He has mentored 1 B.S., 26 M.S., and 5 Ph.D. thesis students; 47 undergraduate research students and 11 undergraduate senior design project teams; over 480 K-12 teachers and 115 high school student researchers; and 18 undergraduate GK-12 Fellows and 59 graduate GK-12 Fellows. Moreover, he directs K-12 education, training, mentoring, and outreach programs that enrich the STEM education of over 1,000 students annually.

Fundamental: Optimizing a Teacher Professional Development Program for Teaching STEM with Robotics Through Design-Based Research

1. Introduction

Prior studies have shown that robotics helps stimulate student excitement and encourage student participation in STEM lessons [1,2]. Furthermore, robotics technology as an educational tool can be applied in the teaching and learning of diverse disciplines such as biological-, chemical-, computer-, medical-, and space sciences; design; engineering; language learning; mathematics; etc. [3-7]. A robotics-focused educational framework offers myriad advantages, for example: i) engaging participants in active learning; ii) providing kinesthetic experiences in teaching and learning; iii) permitting students to visualize and comprehend abstract content knowledge in a concrete and tangible manner; iv) motivating learners intrinsically and extrinsically; and v) improving the overall learning environment, learning satisfaction, learning of disciplinary content, and learning/teaching outcomes [8,9]. Considering these advantages of robotics-aided teaching and learning, application of robotics in STEM education has gained intense interest from educators, policy makers, and teaching administrators, become an area of active research and development, and attracted concerned educators and policy makers to incorporate robotics into STEM curricula [10]. Integration of robotics technology in STEM education as a pedagogical tool has also proved its value in enhancing student engagement and learning outcomes in STEM disciplines.

Nonetheless, robotics-based STEM curricula in middle schools have not received sufficient attention yet [1]. Use of robotics in middle school STEM education (mainly in math and science lessons) is particularly important since children in this age-group start making choices about courses that they are interested to pursue in future [11-15]. The kinesthetic learning and intrinsic and extrinsic motivations for middle school students are also important factors due to their age and maturity levels. Thus, it is critical that teachers engage middle school students in STEM disciplines effectively through the applications of robotics-aided lessons. However, robotics-aided STEM lessons have not received priority yet and such lessons are still not a part of regular curricula at middle schools [1].

To promote the implementation of robotics-aided STEM lessons in middle schools and their incorporation in regular curricula, we posit that it is of paramount importance that the prospective teachers be identified first and then they be trained professionally before being tasked to implement robotics-based lessons in classrooms. However, professional development (PD) programs to train teachers about how to develop and implement such robotics-aided lessons in classrooms are relatively rare, perhaps due to the lack of required resources and facilities. Furthermore, such PD programs should be properly managed and formally evaluated to ensure their effectiveness in meeting the expectations of participants. The programs should also produce appropriate outcomes and prepare the participants in such ways that they gain the maximum knowledge and skills. However, recent literature does not reveal significant number of PD programs for middle school teachers implementing robotics-aided STEM lessons, except a few preliminary initiatives [16,17]. In addition, the state-of-the-art management styles of the PD programs cannot guarantee that the PD programs meet the program objectives making the participants fully prepared for the targeted robotics-aided lessons in actual classroom settings [16]. We argue that prevailing limitations of the few state-of-the-art PD programs for robotics-aided STEM lessons for middle school educators are due to the lack of proper management approaches, program evaluation methods, and continual improvement techniques.

Evan as recent initiatives [16] have attempted to properly manage and evaluate PD programs for middle school teachers for robotics-aided STEM lessons, they have lacked holistic and comprehensive evaluation methods. While qualitative evaluation of such PD programs is important, quantitative evaluation is also necessary. We posit that a balanced combination of qualitative and quantitative evaluation can yield a superior approach to assess the quality of PD programs for middle school teachers for robotics-aided STEM lessons. However, combined evaluation methods, comprising qualitative and quantitative evaluations, of such PD programs showing whether or not the PD programs are effective have not been addressed adequately in literature. We also argue that occasional, one-shot evaluation of such PD programs may not be sufficient to ensure desired outcomes of the PD programs. Based on our recent experiences, we posit that design-based research (DBR) [18] in conjunction with combined qualitative and quantitative evaluation methods can be an effective tool to evaluate and improve PD programs for middle school teachers for robotics-aided STEM lessons, and to ensure desired program outcomes. The DBR method [18] breaks the entire program into several segments, evaluates outcomes of each segment, takes necessary actions to improve the program for the next segments, etc. In addition, quantitative evaluation of PD program performance can also help the DBR method to quantify the prevailing situations and pinpoint necessary actions for improvement in the next iterations. However, DBR methods in conjunction with combined qualitative and quantitative evaluation methods have not been integrated yet to ensure highest level of performance of PD programs for middle school teachers for robotics-aided STEM lessons.

Based on the above rationale, this paper proposes an integrated method to measure and improve the design and implementation performance of a PD program for in-service, middle school teachers on teaching STEM with robotics through DBR. We recruited 23 middle school science and math teachers and designed a three-week, eight-hours per day, PD program for the participants to learn how to develop robotics-aided lessons and implement the developed lessons in classroom environment. Under the iterative refinement framework of DBR [18], we divided the three-week PD program into three successive iterations, i.e., each week was considered as an iteration. We expressed the quality (performance level) of the design and implementation of the program using several key evaluation criteria (key performance indicators or KPIs). We developed an objective function based on the KPIs and assigned weight (importance) to each criterion. We used two separate Likert scales with scores between 1 to 10 to determine the weights and assess the performance level, respectively, of each criterion through teacher surveys. We conducted the surveys at the end of each week (iteration), determined the objective function value, analyzed the outcomes, and took necessary actions to enhance the objective function value in the next iteration(s). Here, the objective function value computed in an iteration indicates the overall performance of that iteration. For the selected KPIs, weights, and scales, the maximum possible objective function value was 1,200. We assume that the objective function value is the targeted performance level, and out of the 3 weeks, the week with the highest objective function value shows the maximum *achieved* performance level. We assume that the constraints were the same/similar in each week. We also adopted two hypotheses related to perceived performance of the PD programs and perceived weight (importance) of each evaluation criteria.

The proposed approach is novel in that it acts as an assessment tool that may bring objectivity in the design and implementation of PD programs. The DBR can act as a continuous improvement tool to enhance the quality of the PD program iteratively. Both approaches can jointly enhance the effectiveness and performance of the PD program.

2. The Professional Development Program

We designed and implemented a PD program for in-service middle school STEM teachers on teaching STEM using robotics. For this purpose, we recruited 23 in-service science and math teachers from New York City middle schools and designed a three-week, eight-hours per day, PD program to teach and train the participants on how to develop robotics-aided lessons and implement the developed lessons in actual classroom environment. We developed a facilitation team that comprised of mechanical engineering graduate students and postdoctoral researchers, education graduate students, engineering professors, education professors/experts, and education administrators. All PD sessions were held at the NYU Tandon School of Engineering. Statistics of the participating teachers and the facilitation team is given in Tables 1 and 2, respectively.

Under the mentorship of engineering and education professors, the graduate students and postdoctoral researchers developed the robotics-aided lessons and other supporting activities and conducted all instructional activities during the PD. In advance of the PD, a three-week schedule was developed to implement the entire program. The program included combination of fundamental educational theories and concepts, robotics fundamentals, and robotics-aided math and science lessons. The project team facilitated the PD through different instructional modes and methods such as lectures, hands-on activities, group discussions, projects, co-generation dialogues, assignments, brainstorming sessions, competitions, challenges, question and answer sessions, etc. An online feedback and reflection system was also implemented to solicit feedback from the middle school teachers at the end of each day.

Table 1: Statistics of the participating teachers.

School/teacher information	Number
Participating schools total	13
Participating teachers total	23
Participating math teachers total/male/female	10/4/6
Participating science teachers total/male/female	13/7/6

Table 2: Statistics of the facilitation team.

Team member category	Number
Engineering graduate students	3
Education graduate students	2
Engineering postdoctoral researchers	2
Engineering professors	2
Education professors	2
Education administrators	1

3. Development of the LEGO Robotic System

We developed a base robot as illustrated in Figure 1 using a LEGO Mindstorms EV3 robotics kit to implement numerous robotics-aided middle school STEM lessons [19]. The developed robotics system included i) a programmable brick that is programmed through a graphical-user interface (GUI), hosts several user-interface buttons and an LCD screen, and that serves as the control unit and power station for the robotic system; ii) two large electric motors to provide powerful and precise action and motion of the robotic vehicle through appropriate program and control; iii) several sensors such as ultrasonic, touch, color, temperature, wheel rotation, gyroscope, etc.; and iv) two wheels, several gears, different types of cables, and various configuration parts and accessories to build the robot structure as needed. We used the LEGO kit being convinced of its relatively affordable cost, simple programming and operation, easy troubleshooting, flexibility in assembly, configuration, and reconfiguration, simple power supply, easy storage, and appropriateness of its functions and capabilities in explaining middle school science and math content [1-2,4,6,8,9,16,19].

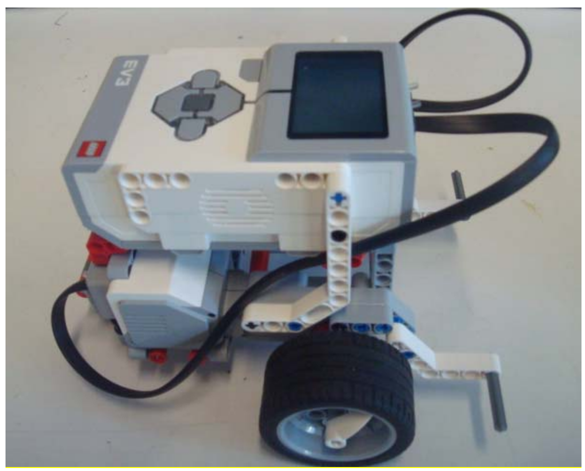

Figure 1: A LEGO Mindstorms EV3 robotic system developed by the teachers to be used in teaching and learning of middle school STEM lessons.

4. The Development of Robot-aided STEM Lessons

The project facilitation team and the PD program participants collaborated in different ways to plan and develop several robotics-aided STEM lessons for middle school classrooms. The teachers shared a few existing middle school science and math concepts that they deemed pedagogically challenging. The research team collaborated with the teachers to develop several interesting and potentially useful robotics-aided lessons, teaching strategies, hands-on activities and activity sheets, lesson description materials, and teaching outcome assessment materials. While doing so, the research team was influenced by myriad relevant education research theories [20-27]. When planning and developing the lessons, it was considered that all lessons must meet the state standards for middle school science and math based on the Common Core State Standards for Math (CCSSM) [28] and the Next Generation Science Standards (NGSS) [29]. Being influenced by DBR, throughout the lesson development and testing process, the project team members and participating teachers performed iterative alterations and modifications to improve the overall quality of the lessons. To do so, we arranged co-generation meetings, brainstorming sessions, group discussions, etc., to upgrade and improve the lessons. Similar activities during the actual PD program further enticed some of the representative and returning teachers to incorporate educational robotic technologies in their advanced lesson plans and helped redesign the lessons based on their local observations, discussions with PD facilitation team, and their existing knowledge of what it would take to actually implement the lessons with students in classroom settings.

It was planned that graduate students and postdoctoral researchers from project facilitation team (Table 2) would visit the collaborating schools during academic year, observe teachers' classroom implementation strategies and techniques of robotics-aided STEM lessons to further justify the effectiveness of the PD program, and establish the fidelity of the lesson implementation. Teachers would also help collect feedback from their students to enhance the lesson contents, pedagogy, and evaluation.

As explained above, we designed several robotics-aided science and math lessons for different middle school grade levels. For examples, the science lessons addressed topics such as mass, force, torque, moment, displacement, energy, environment, velocity, speed, acceleration, gravity, friction, design, design optimization, biology, cell division, biological adaptation, osmosis and diffusion, etc. The math lessons addressed topics such as ratio and proportion, number line, function, analyzing and interpreting data, least common multiple, statistics, expressions and equations, etc. The project team and participating teachers designed and developed necessary attachments and accessories to the robot base of Figure 1, developed new computer programs or modified existing computer programs for the corresponding lessons, and prepared appropriate activity sheets before implementing the lessons in the training sessions during the PD program and later in actual classrooms with students. The trained middle school teachers guided their students to implement the activities using robots during actual class periods, and the students also needed to record the observations in supplied activity sheets. Table 3 provides a brief overview of a representative math lesson and Figure 2 shows a representative picture during the lesson implementation.

Table 3: Description of a representative math lesson.

Lesson topic	Lesson description
Number line	A number line is created on the classroom floor using tape and papers. There is a zero (0) location. Then, it can be repeatedly increased by 1 toward positive direction, and repeatedly decreased by 1 toward negative direction. The robot vehicle can travel from one digit location to another digit location on the number line. Students can learn addition, subtraction, etc., through the movement of the robotic vehicle. Of course, the lesson meets the Common Core State Standards for Math (CCSSM) [28].

Figure 2: The number line lesson being implemented using the LEGO robotic vehicle.

5. Developing the Computational Model of Program Performance

Based on our experiences and brainstorming results with facilitators, education experts, and participating teachers, we opted to express the quality (performance level) of the design and implementation of the program using several KPIs. These KPIs are introduced below with abbreviations in parentheses.

i. Instruction: Quality of instruction by the instructors, i.e., the instruction quality (I_q)
ii. Materials: Quality of instruction materials, i.e., the materials quality (M_q)
iii. Methods: Appropriateness of selection of instruction modes/methods (M_a)
iv. Scheduling: Management of scheduling and timing (M_{st})
v. Equity: Ability to understand diversity, individual differences, and needs to ensure equity (A_{dde})
vi. Contents: Ability to maintain balance in delivered contents, i.e., appropriate combination of math, science, education, and robotics topics (B_{dc})
vii. Prerequisites: Maintaining prerequisites in the delivered topics (M_p)
viii. Feedbacks: Quality of reflections and addressing feedbacks from participants (R_q)
ix. Surveys: Quality of surveys and assessments (A_q)
x. Teaming: Team building and teaming ability (A_{tb})
xi. Innovation: Providing proper guideline toward innovation in new lesson planning (G_{lp})
xii. Engagement: Level of engagement of the participating teachers (L_{ta}).

We developed an objective function as in (1) based on the above criteria and assigned weight (importance) to each criterion. In (1), V_{of} is the computed objective function value for an iteration (week); n is the number of program participants in that week/iteration; MW_{I_q}, $MW_{M_q}, \dots, MW_{L_{ta}}$ are the means of $W_{I_{q_i}}, W_{M_{q_i}}, \dots, W_{L_{ta_i}}$ computed following the example shown in (2) for that week/iteration. As exemplified in (2), $W_{I_{q_i}}, W_{M_{q_i}}, \dots, W_{L_{ta_i}}$ are the weights of I_q, $M_q, \dots, L_{ta}$, respectively, assessed by each program participant in that week/iteration.

$$V_{of} = MW_{I_q} \sum_{i=1}^{n} I_{q_i} + MW_{M_q} \sum_{i=1}^{n} M_{q_i} + \cdots + MW_{G_{lp}} \sum_{i=1}^{n} G_{lp_i} + MW_{L_{ta}} \sum_{i=1}^{n} L_{ta_i} \quad (1)$$

$$MW_{I_q} = \frac{1}{n} \sum_{i=1}^{n} W_{I_{q_i}} \quad (2)$$

We developed two separate Likert scales with scores between 1 to 10 to determine the weight of each criterion ($W_{I_{q_i}}, W_{M_{q_i}}, \dots, W_{L_{ta_i}}$), and quantitatively assess the performance level of each criterion (I_q, $M_q, \dots, L_{ta}$) through teacher-surveys. These scales are given in Appendices A and B, respectively.

6. Iterative (Weekly) Evaluation and Analyses of the Program Performance and Improvement through DBR

We conducted the surveys at the end of each week (iteration), determined the computed objective function value (V_{of}), analyzed the outcomes, and took necessary actions to enhance the V_{of} in the next iteration(s).

For the selected KPIs, weights, and scales, the maximum possible V_{of} (V_{ofmax}) was 1,200. We assume that V_{ofmax} is the targeted performance level, and out of the 3 weeks, the week with the highest V_{of} shows the achieved performance level. We assume that the constraints are the same/similar in each week and are as follows.

1. We cannot add or replace the instructors
2. We cannot arrange separate training program for individual educators during the PD program
3. We cannot make significant changes in the developed lessons, lesson materials, activities, and activity sheets
4. We cannot increase/decrease the training period and change the training dates
5. We cannot stop or postpone the training program temporarily due to potential mismanagement, etc.

We adopt the following two hypotheses:

***Hypothesis 1*:** The DBR process causes the V_{of} to increase (i.e., with successive refinements, the teachers perceive the program to improve) toward the targeted value as the iteration propagates.

***Hypothesis 2*:** The weights of the performance criteria as perceived and evaluated by the teachers vary between iterations.

As mentioned earlier, we determined the computed objective function value (V_{of}) of performance for each week/iteration, analyzed our observations and feedbacks of the participants, and implemented necessary corrective actions to improve the computed objective function value (V_{of}) of performance in the subsequent iterations. For example, the computed objective function value (V_{of}) of performance for the three (3) different weeks (iterations) of the PD program are shown at a glance in Figure 3 (quantitative evaluation results). Here, the computed objective function values (V_{of}) in a week indicates the overall performance of the PD program in that week. Table 4 shows, as an example, the qualitative feedbacks/comments we received from the program participants regarding their evaluation of the overall program in terms of the proposed 12 performance criteria at the end of week 1 (iteration 1), the qualitative observations we made ourselves on the 12 performance criteria in week 1, and the corrective actions we took on the 12 performance criteria to improve the performance of the PD program in week 2 based on the evaluations from week 1. In Table 4, the digits inside parentheses indicate the frequency of similar feedback provided by PD program participants. The higher frequency means that the feedback is more important and it should be prioritized when making improvement decisions. We performed similar analyses for the remaining iterations (2 and 3).

The results in Figure 3 show that the performance of the PD program gradually improved from iteration to iteration due to the implementation of the DBR method. It is true that the amount of improvement in the computed objective function values (V_{of}) between the iterations is not very high, but it still shows improvement, which may be sufficient to prove the effectiveness of the proposed DBR method. In fact, the effectiveness of the DBR depends on how the observations of one iteration are analyzed and how corrective actions are adopted for next iteration(s). The proposed DBR method might result in higher amount of improvement from iteration to iteration if the observations of one iteration could be analyzed and corrective actions could be adopted for next iteration(s) in better ways. The results in Figure 3 justify Hypothesis 1 that the DBR process causes the V_{of} to increase (i.e., with successive refinements, the teachers perceive the program to improve) toward the targeted value as the iteration propagates.

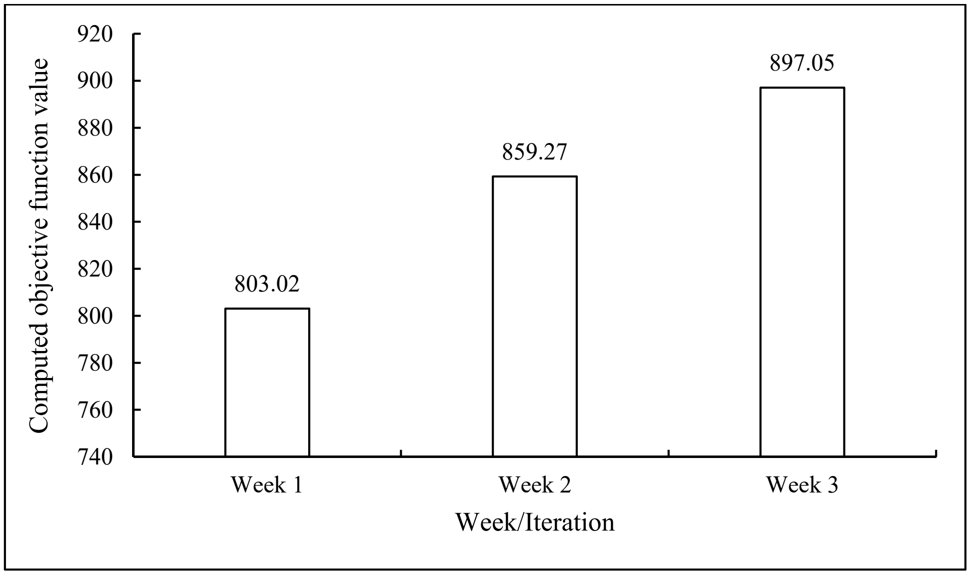

Figure 3: Computed objective function values (V_{of}) of performance for different weeks (iterations) of the PD program.

Table 4: Qualitative feedbacks from participants and qualitative observations of the facilitation team on the 12 performance criteria, and the corrective actions adopted by the facilitation/research team on the 12 performance criteria at the end of week 1.

Performance Criteria	Feedbacks from PD participants	Observations of the facilitation team (researchers)	Corrective actions taken for next iteration (s)
Instruction	Instructions on some portion of hands-on activities were unclear. There was noise from other participants and some instructors could not help reduce it, thus restricting understanding of the instructions. Detailed explanation on sensor use was necessary. Sometimes instructors wrote too much on the board and participants needed to take too much notes, but note taking wasn't really an option in that setting. Robot programming was pre-prepared, it would be easier to understand if the presenter/instructor coded the program with the participants. Directions on hands-on robot activities provided by instructors were difficult to follow.	Some instructors repeated the same sentences several times. Start and end of the instruction were not smooth. Pronunciation and accent were not appropriate. Speed of lecture was not uniform and well-adjusted with the need of majority participants. Instructors and participants did not proceed with same pace. Some questions raised by participants were not responded properly. Individual care was needed, but not provided properly. Overall presentation was not well-prepared; eye-contact between instructors and participants were not appropriate, etc.	We read the feedback of the participants at the end of the day, talked with all concerned instructors immediately after that about the issues raised by PD participants, and asked the instructors to improve on the identified matters. We also observed each instructor when he/she instructed and notified him/her at the end of the session, if we noticed any issue.

Materials	Some outline materials at the beginning of many topics seem to be unnecessary. Some portions of the activity sheets were less necessary to understand the point and thus could be removed. The lecture materials could be given before the lectures not after the lectures that seem to be necessary for better understanding. The robotics kits could be mailed to schools directly instead of distributing every day. Some of the lessons were less meaningful because of the functionalities (the biological adaptation of the robots trying to go through snow but have ineffective legs). There should have been more number lines on the floor so more of the participants could test the lesson at the same time. Chart paper should be provided during small group discussions so the participants could collectively brainstorm and record their ideas as a group. A handout that explains all the components of the programs could be supplied.	Some materials were not updated properly. Some parts of materials were not used in the class or in hand-on activities, but were not removed. Sometimes there were duplicate copies of the same materials of different versions that caused problems. Some printed materials were not legible, some soft materials were not well-explained. Some materials explained allied concepts of the main topics, but the participants were not interested in those allied and redundant topics, etc.	We read the feedback of the participants at the end of the day, talked with all concerned instructors immediately after that about the issues related to instruction materials raised by PD participants, and asked the instructors to solve the problems with the materials concerned. We also observed each instructor when he/she instructed, and notified him/her at the end of the session, if we noticed any issue related to instruction materials.
Methods	The instructions were sometimes one-sided and opportunity to ask questions to the instructors were sometime less. Lecture-style sessions should be reduced and more discussions and brainstorming sessions should be added as most of the participants were experienced and they also had something to add. One-on-one questioning, individual help, and feedback is necessary particularly during programming lectures (3). Inquiry-based teaching styles were necessary, but it was not so prioritized. It would have been nice to discuss at least two of the questions from the worksheet so that participants who do not teach science could get a much deeper understanding of how science related lessons work (e.g., center of mass relates to gravity). This could possibly help some participants, for example, math concept that can be linked to a particular science for cross curriculum lesson. The program was productive but often the lesson was not really appropriate for the skill being introduced. It would be much more productive if the program progressed at a “go at your own speed” type of pace.	Sometimes the instructors followed lecture-style teaching, but the participants were expecting discussion or mixture of lectures and discussions. Involvement of the participants was less during the instructions.	We reviewed the feedback of the participants at the end of the day, talked with all concerned instructors immediately after that about the issues raised by the PD participants, and asked the instructors to improve the methods based on participant’s opinions. We also observed each instructor when he/she instructed, and notified him/her at the end of the session, if we noticed any issue related to methods of instructions.

	It would also have been preferred to have more actual discussions within the sections of the day labeled "Discussions" than actually happened. There was really no incentive to speak to neighbors or collaborate with them on code/troubleshooting. More interaction was needed with other participants especially during technical sessions (2). The lesson should have been presented as if participants were the students, making it easier to understand the goal of the lesson, but it was not exactly the same.		
Scheduling	The schedule was tight and many topics were accommodated in compacted timing, more time was needed to finish some topics (3). There were a few unnecessary breaks between sessions, some sessions were given too long unnecessary time. There were many downtimes/waiting times and it was unproductive and the teams that completed tasks could be allowed to proceed further (6). Feedback session was too short to write feedback clearly. Time to solve the challenge problems seems to be insufficient (3). A more realistic amount of time to complete tasks during instructional portions should be provided. Participants would have liked more time spent on the programming (mathematical operations). Discussion materials could be given many days before the sessions so that participants could read those beforehand (2). More time should be allotted to manage contingency as sometimes robot may need troubleshooting. Only programming topics but almost no robotics-aided lessons in the first week.	Sometimes sessions started late due to lack of attendance of the participants, which also affected other subsequent events in the schedule. Some sessions took significantly more time that caused bored situations. Sometimes, sessions were finished very quickly without giving sufficient time to understand and practice the lessons. Sometimes schedule was changed for unanticipated reasons and was notified to the participants just before the activities.	We reviewed the feedback of the participants at the end of the day, talked with all concerned instructors immediately after that about the issues raised by the PD participants, and asked the instructors to improve them on the identified matters of maintaining timing and schedule. We also observed each instructor when he/she instructed and notified him/her at the end of the session, if we noticed any issue related to schedule and time. From PD management point of view, the program manager thought critically to set program events in the revised schedule so that minimal changes in the schedule could be achieved.
Equity	Some participants were slow in understanding and some were fast, but the slow participants required more supports from the instructors. There was a mixture of science and math topics, and hence the science teachers could not enjoy the math lessons as those were enjoyed by math teachers, and the same happened to the math teachers; etc.	Instructors were not very attentive to take individual care of the participants. Slow learners felt helpless and could not proceed properly. Some participants used seats that were very close to the instruction table, others were far and they might not be able to listen equally and they might also fail to draw the attention of the instructors.	We reviewed the feedback of the participants at the end of the day, talked with all concerned instructors immediately after that about the issues raised by the PD participants if any of the participants could not receive benefits equally, and asked the instructors to improve the methods and overall instructions based on participants' opinions to

			ensure equity. We also observed each instructor when he/she instructed, and notified him/her at the end of the session, if we noticed any issue related to equity.
Contents	The programming load was very high (2). The contents could more focus on standards. The morning session on the first day was a bit boring because of the paperwork aspect of the presentation. Programming challenge with the color sensor was very frustrating because the robot was not detecting the colors accurately. Programming was complex and challenging (2).	Too much programming was put at the beginning part of the PD program that conveyed a misconception to the participants about the objective of the PD program. Proper combination of robotics, education, and general subjects was not so optimized in the first week.	We reviewed the feedback of the participants at the end of the day, talked with all concerned instructors immediately after that about the issues raised by the PD participants, and asked the instructors to improve the session contents based on participants' opinions. We also observed each instructor when he/she instructed, and notified him/her at the end of the session if we noticed any issue related to lecture contents. From the PD management point of view, the program manager thought critically to determine contents of each session/event so that the overall content looked balanced.
Prerequisites	It was slightly difficult to follow the instructions as some technical aspects were unknown previously and were taught for the first time. Participants got a bit frustrated when doing the surveys and the questions were being asked as if they already knew how to use robotics (2). Surveys were difficult to follow as many questions were such that participants did not feel ready to answer them. Some parts of surveys were not applicable at that stage of the PD program.	Some topics were discussed before the pre-requisites were taught. For example, some lessons were taught before teaching applications of some programming blocks and some sensor applications. Some education terminologies were put in the lesson description that were not explained properly in prior classes.	We reviewed the feedback of the participants at the end of the day, talked with all concerned instructors immediately after that about the issues raised by the PD participants, and asked the instructors to improve the session contents based on prerequisites. We also observed each instructor when he/she instructed, and notified him/her at the end of the session if we noticed any issue related to pre-requisites of lecture contents. From the PD management point of view, the program manager thought critically to determine contents of each session/event so that the overall contents looked

			balanced and prerequisites are maintained.
Feedbacks	There was too much noise during feedback sessions.	Feedback sessions were short as the participants were not interested to spend much time to write feedbacks at the end of the day. Importance of feedback was not so clear to the participants. Some participants neglected to write feedbacks timely or they did not write the feedbacks at all.	We reviewed the feedback of the participants at the end of the day, talked with all concerned instructors immediately after that about the issues raised by the PD participants, and asked the instructors to improve their services based on the feedback. From the PD management point of view, the program manager thought critically to administer the feedback sessions so that the participants could provide feedback appropriately.
Surveys	Surveys were unclear as many topics were unknown. Surveys were a bit overwhelming (5). Surveys were a lot, somewhat repetitive (4). Some of the surveys did not make sense (2). Surveys could be more engaging, for example, a live survey where everyone answers each question from their laptops while the results are displayed on the board would be fun. Perhaps surveys could be included in the application process or at least could be e-mailed to the participants before the day so that they wouldn't need to spend the entire morning completing them (2). Surveys and pre-assessments were difficult. Some surveys were online that were not good.	Some participants felt bored with surveys and asked to reduce the survey volume and frequency. Some participants asked to know whether they could take the surveys at home and at later times. Some survey questions were difficult to understand. Objective of the surveys might not be clear to the participants. Some surveys were repeated but the participants could not understand its reasoning and thus complained.	We reviewed the feedback of the participants at the end of the day, talked with all concerned instructors immediately after that about the issues raised by the PD participants, and tried to reduce the loads of surveys, explain the surveys, and give them flexibility to respond to the surveys at home if feasible. From the PD management point of view, the program manager thought critically to determine contents of each survey session so that the overall contents looked balanced and participants did not feel bored.
Teaming	When teaming for hands-on activities, it would be better if the teams were formed with teachers of same subjects and grades, sometimes it was not followed (3). Perhaps the math teachers could show the science teachers what they brainstormed during separate discussions as such communication between separate teams is necessary (2).	Sometimes participants were absent or arrived late that caused problem in teaming. Sometimes there was lack of understanding between team members. Each team was formed with one math and one science teacher, and thus both teachers did not concentrate on both math and science related topics and lessons.	We reviewed the feedback of the participants at the end of the day, talked with all concerned instructors immediately after that about the issues raised by the PD participants, and asked the instructors to improve team spirit among teachers while instructing. From the PD management point of view, the program manager thought critically to determine appropriate

			team members and change team members if necessary.
Innovation	Participants needed to be too much busy with regular topics and activities, and emphasis on reasoning, thinking, innovation was less. There is no clear guideline to perform more advanced and novel programming with different unseen problems. Discussion on alternative ideas to improve the presented lessons was less.	Instructors were busy to teach regular contents and discussion/brainstorming was not adequate to stimulate innovation in participants. Dedicated time to practice innovation was also not available.	We reviewed the feedback of the participants at the end of the day, talked with all concerned instructors immediately after that about the issues raised by the PD participants, and asked the instructors to improve the innovative ideas. From the PD management point of view, the program manager thought critically to improve innovation especially during new lesson planning and development.
Engagement	There were too many agenda items and activities that caused fatigue and reduced engagement.	Some participants talked among themselves unnecessarily. Preliminary work sampling study showed that the participants were not equally concentrated in the activities during the class period, instead they were busy with some other side works such as talking over phones, browsing internet, writing or reading personal notes, etc. Some participants passed their time without doing anything.	We reviewed the feedback of the participants at the end of the day, talked with all concerned instructors immediately after that about the issues raised by the PD participants, and asked the instructors to improve the engagement of the participants. From the PD management point of view, the program manager thought critically to determine the activities of each session/event so that the teachers needed to be engaged with the activities by default.

To further analyze the results presented in Figure 3, we show the mean performance for different performance criteria (KPIs) for 3 different weeks (iterations) of the PD program in details as in Figure 4. The results show that performance of individual criterion improved from week 1 to week 2 except a few criteria such as materials quality, equity, and balance in PD program contents. Similarly, performance of individual criterion improved from week 2 to week 3 except a few criteria such as materials quality, scheduling and timing, equity, balance in PD program contents, and survey and feedback quality. The results show that the ratings of materials, instruction methods, equity, content, survey, and engagement did

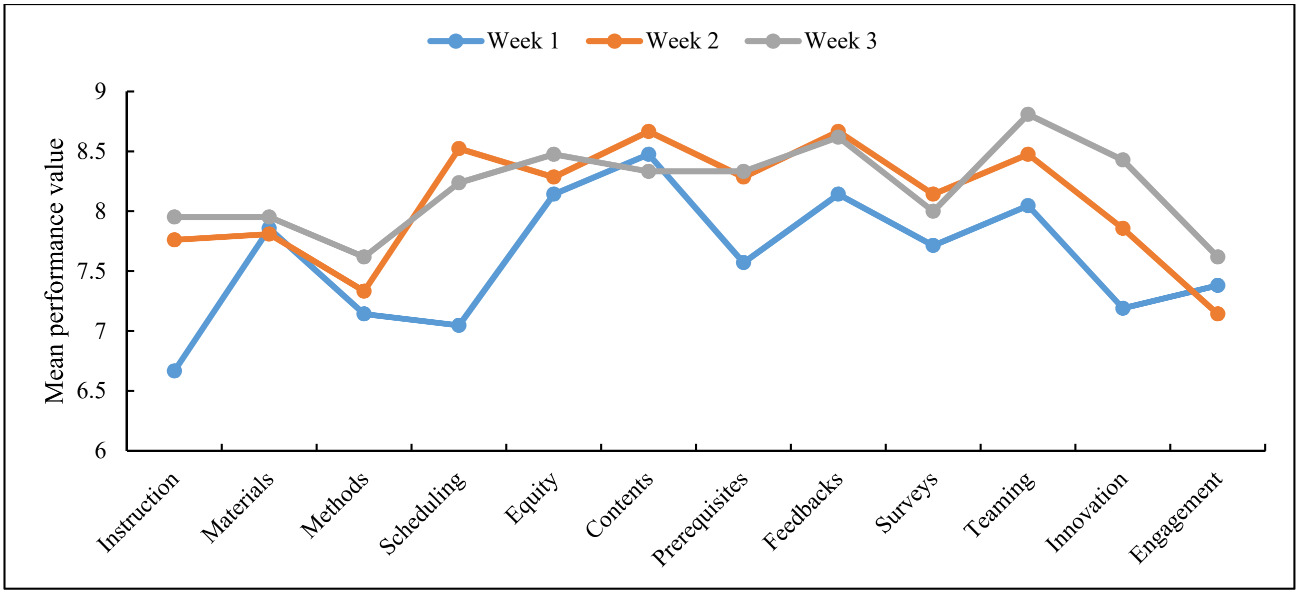

Figure 4: Mean performance for different performance criteria (KPIs) for 3 different weeks (iterations) of the PD program.

not improve much even after the 3rd (the last) week/iteration. Results of paired *t*-tests for variations in participants' perceived performance criteria of the PD program among the weeks/iterations are shown in Table 5. The results show that variations in performance criteria between iterations are significant for two out of three cases. We believe that such variations caused the improvement in terms of computed objective function values through the DBR as shown in Figure 3. However, it would be better if the variations in performance criteria between iterations for the third case could also be significant. The variations in performance criteria between iterations were not significant or were less significant, which might happen due to the reasons that observations on these criteria in one iteration were not analyzed properly or the analyses were performed properly but appropriate corrective actions were not adopted and implemented during the subsequent iteration(s) to produce significant improvement (changes/variations) in the PD program performance. It can also mean that the existing constraints did not allow the facilitation team to make necessary improvement/changes beyond a certain level. However, the variations were sufficient to partly justify Hypothesis 1 that the DBR can improve the PD program performance [18].

Mean weight values for different performance criteria (KPIs) perceived by the participants at three different weeks (iterations) of the PD program are shown in Figure 5. To examine the second hypothesis, variations in weights perceived by the participating teachers between the iterations are analyzed. Results of paired *t*-tests for variations in participants' perceived weights of the performance criteria among the weeks/iterations are shown in Table 5. Variations in perceived weights by participants between weeks/iterations were significant for two out of three cases, partly justifying Hypothesis 2. The results thus show that the weights (importance) of different performance criteria (KPIs) are perceived differently by the participants for different consecutive weeks/iterations. It means that the participants expected different levels of performance of all 12 performance criteria throughout the entire PD program. In general, the perceived weights of the performance criteria increase as the iterations propagate. It means that the participants realize the importance of the performance criteria for overall success of the PD program more as they gain more experiences of the PD program.

Table 5: Results of paired *t*-tests for variations in participants' perceived performance of the PD program and perceived weights of the performance criteria among the weeks/iterations

Item	Domain	*t* value	*p* value	Significance
Performance	Week 1 v/s Week 2	2.99	0.0030	Yes
	Week 2 v/s Week 3	1.03	0.3044	No
	Week 3 v/s Week 1	3.65	0.0003	Yes
Weight	Week 1 v/s Week 2	0.6954	0.4875	No
	Week 2 v/s Week 3	2.4681	0.0143	Yes
	Week 3 v/s Week 1	2.7803	0.0058	Yes

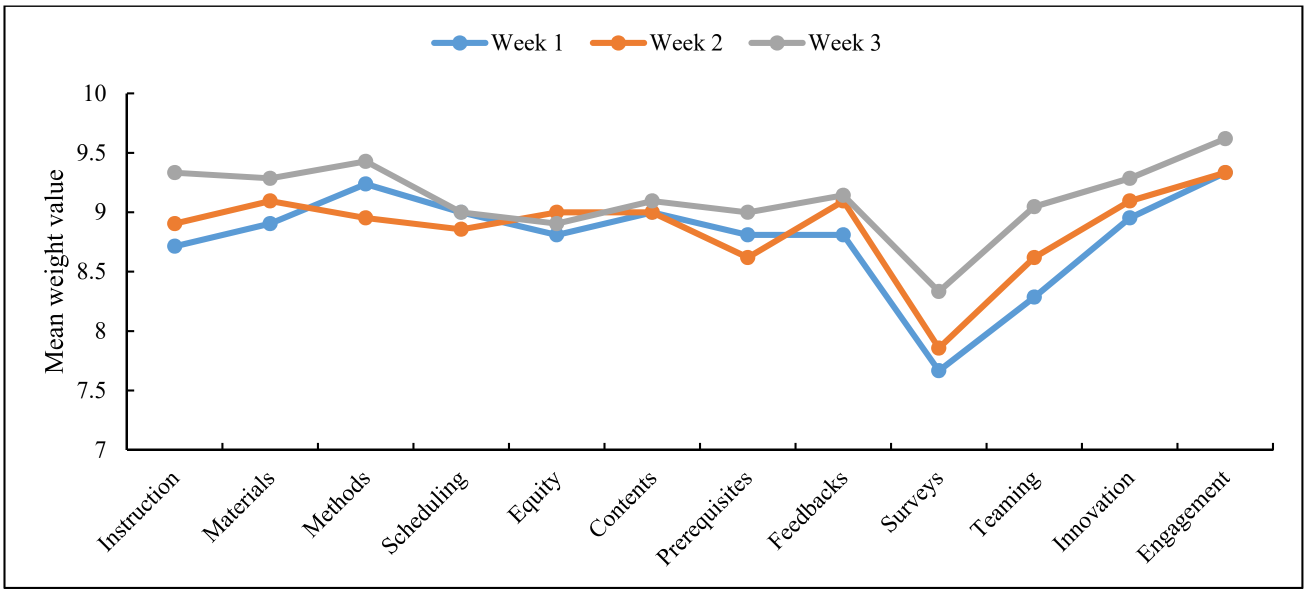

Figure 5: Mean weight values for different performance criteria (KPIs) perceived by participants at 3 different weeks (iterations) of the PD program.

The results in Figure 5 show that the mean perceived weight values for all criteria are high (above 8 out of 10) except for surveys, which indicates that:

1. The performance criteria of the PD proposed by us for our performance computational model proposed in equation (1) were also perceived as important by the participants for successful implementation of the PD program, which validates the proposed model and that the model does not include any meaningless or less important criteria.
2. The participants were not too much interested in the surveys or those were overwhelming. It might happen because the participants might think that the surveys were not the necessary parts of their lessons, or they were bored with the extensive volume of surveys. However, the surveys were very important for the researchers to manage and improve the performance of the overall PD program.

7. Conclusions and Future Work

We implemented a method, through DBR, to measure and improve the design and implementation performance of a PD program for in-service teachers on teaching STEM with robotics. We recruited middle school science and math teachers and designed a three-week PD program for the participants to enable them to learn how to develop robotics-aided lessons and implement the developed lessons in classroom settings. Under the iterative refinement of a framework called the DBR, we divided the three-week PD program into three successive iterations/weeks, i.e., each week was considered as an iteration. We selected a few factors to express the quality (performance level) of the design and implementation of the program (KPIs). We developed an objective function based on those criteria and assigned weight to each criterion. The objective function value for a week/iteration indicates the overall performance level of the entire PD program for that week/iteration. The results show that the proposed integrated method can evaluate the PD program quantitatively and improve its performance through iterations. The proposed approach is novel in the sense that it acts as an assessment tool to bring objectivity in the design, implementation, and evaluation of PD program. The DBR acts as a continuous improvement tool to enhance the quality of the program iteratively. Both approaches jointly enhance the effectiveness and performance of the PD program. The methods also include qualitative assessment and analysis. The proposed integrated approach (quantitative and qualitative evaluations and DBR) can be used to enhance the effectiveness of PD and educational programs, in general.

In future, we will identify additional performance criteria to assess the PD program. We will use a systems approach to analyze the DBR outcomes. We will use better optimization methods to optimize the overall performance of the PD program. We will develop an online software system so that the participants can evaluate the performance and weights online, and the results (graphs) are automatically prepared and displayed in real-time showing the overall improvements in the performance of the PD program between iterations. The integrated, flexible, real-time program assessment and improvement online software will be commercialized fostering entrepreneurship in educational research and development.

Acknowledgements

This work is supported in part by the National Science Foundation grants DRK-12 DRL: 1417769, ITEST DRL: 1614085, and RET Site EEC: 1542286, and NY Space Grant Consortium grant 76156-10488. The authors thank the middle school teachers for their participation in this study.

References

1. Rahman, S.M.M., Chacko, S.M., and Kapila, V. "Building trust in robots in robotics-focused STEM education under TPACK framework in middle schools," *Proc. ASEE Annual Conference* (2017): https://peer.asee.org/27990.
2. Mosley, P. and Kline, R. "Engaging students: A framework using LEGO robotics to teach problem solving." *Information Technology, Learning, and Performance Journal* 24 (2006): 39-45.
3. Chen, N.S., Quadir, B., and Teng, D.C. "Integrating Book, Digital Content and Robot for Enhancing Elementary School Students' Learning of English." *Australasian Journal of Educational Technology* 27.3 (2011): 546-561.
4. Cruz-Martín, A., *et al.* "A LEGO Mindstorms NXT approach for teaching at data acquisition, control systems engineering and real-time systems undergraduate courses." *Computers & Education* 59.3 (2012): 974-988.

5. Kunkler, K. "The role of medical simulation: An overview." *The International Journal of Medical Robotics and Computer Assisted Surgery* 2.3 (2006): 203-210.
6. Lawhead, P. B., et al. "A road map for teaching introductory programming using LEGO Mindstorms robots." *ACM SIGCSE Bulletin* 35.2 (2003): 191-201.
7. Pomalaza-Raez, C. and Groff, B.H. "Retention 101: Where robots go…students follow." *Journal of Engineering Education* 92.1 (2003): 85-90.
8. Panadero, C.F., Romá, J.V., and Kloos, C.D. "Impact of learning experiences using LEGO Mindstorms in engineering courses." *Proc. IEEE Education Engineering* (*EDUCON*), (2010): 503-512.
9. Whitman, L., and Witherspoon, T. "Using LEGOs to interest high school students and improve K12 STEM education." *Proc. ASEE/IEEE Frontiers in Education Conference*, (2003): p.F3A6-10.
10. Rahman, S.M.M., Krishnan, J.V., and Kapila, V. "Exploring the dynamic nature of TPACK framework in teaching STEM using robotics in middle school classrooms," *Proc. ASEE Annual Conference and Exposition* (2017): https://peer.asee.org/28336.
11. Beaton, A.E., *et al. Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study* Chestnut Hill, MA: Boston College. (1996).
12. Cummings, S., and Tabebel, D. "Sexual inequality and the reproduction of consciousness: An analysis of sex-role stereotyping among children." *Sex Roles* 6.4 (1980): 631-644.
13. Eichinger, J. "Successful students' perceptions of secondary school science." *School Science and Mathematics* 97.3 (1997): 122–131.
14. Gibson, H.L. and Chase, C. "Longitudinal impact of an inquiry-based science program on middle school students' attitudes toward science." *Science Education* 86.5 (2002): 693-705.
15. Silverman, S. and Pritchard, A.M. "Building their future: Girls and technology education in Connecticut." *Journal of Technology Education*. 7 (1996): 41-54.
16. Moorhead, M., Elliott, C.H., Listman, J.B., Milne, C.E., and Kapila, V. "Professional Development through Situated Learning Techniques Adapted with Design-Based Research." *Proc. ASEE Annual Conference* (2016): 10.18260/p.25967.
17. You, H.S., and Kapila, V. "Effectiveness of professional development: Integration of educational robotics into science and math Curricula," *Proc. ASEE Annual Conference* (2017): https://peer.asee.org/28207.
18. Barab, S., and Squire, K. "Design-based research: Putting a stake in the ground." *The Journal of the Learning Sciences* 13.1 (2004): 1-14.
19. https://education.lego.com/en-us (2017).
20. Blumenfeld, P.C., *et al.* "Motivating project-based learning: Sustaining the doing, supporting the learning." *Educational Psychologist* 26.3-4 (1991): 369-398.
21. Bransford, J.D., *et al.* "Anchored instruction: Why we need it and how technology can help." *Cognition, Education, and Multimedia: Exploring Ideas in High Technology,* (1990): 115-141.
22. Brown, J.S., Collins, A., and Duguid, P. "Situated cognition and the culture of learning." *Educational Researcher* 18.1(1989): 32-42.
23. Collins, A., Brown, J.S., and Newman, S.E. "Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics." *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser* (Resnick, L.G. (ed.)) (1989): 453-494.
24. Lave, J., and Wenger, E. *Situated learning: Legitimate peripheral participation*. Cambridge University Press, (1991).
25. Ryan, R.M., and Deci, E.L. "Intrinsic and extrinsic motivations: Classic definitions and new directions." *Contemporary Educational Psychology* 25.1 (2000): 54-67.
26. Savery, J.R., and Duffy, T.M. "Problem based learning: An instructional model and its constructivist framework." *Educational Technology* 35.5 (1995): 31-38.
27. The Cognition and Technology Group at Vanderbilt. "Anchored instruction and its relationship to situated cognition." *Educational Researcher* 19.6 (1990): 2-10.
28. CCSSM. *Common Core State Standards for Mathematics. Common Core Standards Initiative*. Online: http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf, (2010).
29. NGSS. *Next Generation Science Standards* (*NGSS*)*: For States, by States*. Washington, DC: The National Academies Press. Online: http://www.nextgenscience.org/, (2013).

Appendix A

Teacher's Subject: Math/Science   Teacher's Gender: Female/Male

Survey on weight (importance) of the performance criteria of the PD program: Week 1/2/3

(1) Please circle a number out of the ten numbers given as the response
(2) Here, 1 indicates extremely *unimportant*, and 10 indicates extremely *important* for the overall success of the PD program

Statement 1: How important it is to maintain very high quality of the instructions of the instructors?

Response: 1 2 3 4 5 6 7 8 9 10

Statement 2: How important it is to maintain very high quality of the instruction materials?

Response: 1 2 3 4 5 6 7 8 9 10

Statement 3: How important it is to select appropriate mode/method of instructions (e.g., lecture, discussion, group discussion, interviews, self-reading, hands-on activity, etc.) for each topic?

Response: 1 2 3 4 5 6 7 8 9 10

Statement 4: How important it is to manage the program schedule and timing properly?

Response: 1 2 3 4 5 6 7 8 9 10

Statement 5: How important it is to understand diversity, individual differences, and needs to ensure equity?

Response: 1 2 3 4 5 6 7 8 9 10

Statement 6: How important it is to maintain a balance in the delivered contents (i.e., to maintain appropriate combination of math, science, education and robotics topics)?

Response: 1 2 3 4 5 6 7 8 9 10

Statement 7: How important it is to maintain the pre-requisites in the sequence of the teaching topics properly?

Response: 1 2 3 4 5 6 7 8 9 10

Statement 8: How important it is to receive and address the reflections and feedbacks of the participants properly?

Response: 1 2 3 4 5 6 7 8 9 10

Statement 9: How important it is to maintain very high quality of the surveys and assessments?

Response: 1 2 3 4 5 6 7 8 9 10

Statement 10: How important it is to build and maintain the teams properly?

Response: 1 2 3 4 5 6 7 8 9 10

Statement 11: How important it is to provide proper guidelines for innovations in the activities?

Response: 1 2 3 4 5 6 7 8 9 10

Statement 12: How important it is to make the events interesting, keep the participants engaged and ensure that they do not feel bored?

Response: 1 2 3 4 5 6 7 8 9 10

Appendix B

Survey on PD program performance: Week 1/2/3

Notes:

(1) Please circle a number out of the ten numbers given as the response
(2) Here, **1 indicates extremely/strongly disagree**, and **10 indicates extremely/strongly agree**

Statement 1: The overall instruction quality of the instructors was very high

Response: 1 2 3 4 5 6 7 8 9 10

Statement 2: The quality of the instruction materials was very high

Response: 1 2 3 4 5 6 7 8 9 10

Statement 3: Selection of the mode/method of instructions (e.g., lecture, discussion, group discussion, interviews, self-reading, hands-on activity, etc.) for each topic was appropriate

Response: 1 2 3 4 5 6 7 8 9 10

Statement 4: Program schedule and timing were managed properly

Response: 1 2 3 4 5 6 7 8 9 10

Statement 5: Diversity, individual differences, equity (i.e., needs of the participants) were addressed properly

Response: 1 2 3 4 5 6 7 8 9 10

Statement 6: There was a balance in the delivered contents (appropriate combination of math, science, education and robotics topics was maintained)

Response: 1 2 3 4 5 6 7 8 9 10

Statement 7: Pre-requisites in the sequence of the teaching topics were maintained properly

Response: 1 2 3 4 5 6 7 8 9 10

Statement 8: Reflections and feedbacks of the participants were received and addressed properly

Response: 1 2 3 4 5 6 7 8 9 10

Statement 9: Quality of the surveys and assessments was very high

Response: 1 2 3 4 5 6 7 8 9 10

Statement 10: Teams were built and maintained properly

Response: 1 2 3 4 5 6 7 8 9 10

Statement 11: There were proper guidelines for innovations in the activities

Response: 1 2 3 4 5 6 7 8 9 10

Statement 12: The events were interesting, the participants were engaged and they did not feel bored

Response: 1 2 3 4 5 6 7 8 9 10