

RESEARCH

Open Access



Training doctoral students in critical thinking and experimental design using problem-based learning

Michael D. Schaller^{1*}, Marieta Gencheva¹, Michael R. Gunther¹ and Scott A. Weed¹

Abstract

Background Traditionally, doctoral student education in the biomedical sciences relies on didactic coursework to build a foundation of scientific knowledge and an apprenticeship model of training in the laboratory of an established investigator. Recent recommendations for revision of graduate training include the utilization of graduate student competencies to assess progress and the introduction of novel curricula focused on development of skills, rather than accumulation of facts. Evidence demonstrates that active learning approaches are effective. Several facets of active learning are components of problem-based learning (PBL), which is a teaching modality where student learning is self-directed toward solving problems in a relevant context. These concepts were combined and incorporated in creating a new introductory graduate course designed to develop scientific skills (student competencies) in matriculating doctoral students using a PBL format.

Methods Evaluation of course effectiveness was measured using the principals of the Kirkpatrick Four Level Model of Evaluation. At the end of each course offering, students completed evaluation surveys on the course and instructors to assess their perceptions of training effectiveness. Pre- and post-tests assessing students' proficiency in experimental design were used to measure student learning.

Results The analysis of the outcomes of the course suggests the training is effective in improving experimental design. The course was well received by the students as measured by student evaluations (Kirkpatrick Model Level 1). Improved scores on post-tests indicate that the students learned from the experience (Kirkpatrick Model Level 2). A template is provided for the implementation of similar courses at other institutions.

Conclusions This problem-based learning course appears effective in training newly matriculated graduate students in the required skills for designing experiments to test specific hypotheses, enhancing student preparation prior to initiation of their dissertation research.

Keywords Graduate, Training, Critical thinking, Experimental design, Problem-based learning, Doctoral Student

*Correspondence:

Michael D. Schaller
mschaller@hsc.wvu.edu

¹Department of Biochemistry and Molecular Medicine, West Virginia University School of Medicine, Robert C. Byrd Health Sciences Center 64 Medical Center Drive, P.O. Box 9142, Morgantown, WV 26506, USA



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

For over a decade there have been calls to reform biomedical graduate education. There are two main problems that led to these recommendations and therefore two different prescriptions to solve these problems. The first major issue is the pursuit of non-traditional (non-academic) careers by doctorates and concerns of adequate training [1, 2]. The underlying factors affecting career outcomes are the number of PhDs produced relative to the number of available academic positions [1, 3–5], and the changing career interests of doctoral students [6–9]. One aspect in the proposed reformation to address this problem is incorporation of broader professional skills training and creating awareness of a greater diversity of careers into the graduate curriculum [1, 4, 5]. The second issue relates to the curricula content and whether content knowledge or critical scientific skills should be the core of the curriculum [10, 11]. The proposed reformation to address this issue is creation of curricula focusing upon scientific skills, e.g. reasoning, experimental design and communication, while simultaneously reducing components of the curricula that build a foundational knowledge base [12, 13]. Components of these two approaches are not mutually exclusive, where incorporation of select specialized expertise in each area has the potential to concurrently address both issues. Here we describe the development, implementation and evaluation of a new problem-based learning (PBL)-based graduate course that provides an initial experience in introducing the scientific career-relevant core competencies of critical thinking and experimental design to incoming biomedical doctoral students. The purpose of this course is to address these issues by creating a vehicle to develop professional skills (communication) and critical scientific skills (critical thinking and experimental design) for first year graduate students.

One approach that prioritizes the aggregate scientific skill set required for adept biomedical doctorates is the development of core competencies for doctoral students [5, 14, 15], akin to set milestones that must be met by medical residents and fellows [16]. Key features of these competencies include general and field-specific scientific knowledge, critical thinking, experimental design, evaluation of outcomes, scientific rigor, ability to work in teams, responsible conduct of research, and effective communication [5, 14, 15]. Such competencies provide clear benchmarks to evaluate the progress of doctoral students' development into an independent scientific professional and preparedness for the next career stage. Historically, graduate programs relied on traditional content-based courses and supervised apprenticeship in the mentor's laboratory to develop such competencies. An alternative to this approach is to modify the graduate student curriculum to provide a foundation for these

competencies early in the curriculum in a more structured way. This would provide a base upon which additional coursework and supervised dissertation research could build to develop competencies in doctoral students.

Analyses of how doctoral students learn scientific skills suggest a threshold model, where different skillsets are mastered (a threshold reached), before subsequent skillsets can be mastered [17, 18]. Skills like using the primary literature, experimental design and placing studies in context are earlier thresholds than identifying alternatives, limitations and data analysis [18]. Timmerman et al. recommend revision of graduate curricula to sequentially build toward these thresholds using evidence-based approaches [18]. Several recent curricular modifications are aligned with these recommendations. One program, as cited above, offers courses to develop critical scientific skills early in the curriculum with content knowledge provided in later courses [12, 13]. A second program has built training in experimental design into the coursework in the first semester of the curriculum. Improvements in students experimental design skills and an increase in self-efficacy in experimental design occurred over the course of the semester [19]. Other programs have introduced exercises into courses and workshops to develop experimental design skills using active learning. One program developed interactive sessions on experimental design, where students give chalk talks about an experimental plan to address a problem related to course content and respond to challenges from their peers [20]. Another program has developed a workshop drawing upon principles from design thinking to build problem solving skills and creativity, and primarily uses active learning and experiential learning approaches [21]. While these programs are well received by students, the outcomes of training have not been reported. Similar undergraduate curricula that utilize literature review with an emphasis on scientific thought and methods report increased performance in critical thinking, scientific reasoning and experimental design [22, 23].

It is notable that the changes these examples incorporate into the curriculum are accompanied with a shift from didactic teaching to active learning. Many studies have demonstrated that active learning is more effective than a conventional didactic curriculum in STEM education [24]. Problem-based learning (PBL) is one active learning platform that the relatively new graduate program at the Van Andel Institute Graduate School utilizes for delivery of the formal curriculum [25]. First developed for medical students [26], the PBL learning approach has been adopted in other educational settings, including K-12 and undergraduate education [27, 28]. A basic tenet of PBL is that student learning is self-directed [26]. Students are tasked to solve an assigned problem and are required to find the information necessary for

the solution (self-directed). In practice, learning occurs in small groups where a faculty facilitator helps guide the students in identifying gaps in knowledge that require additional study [29]. As such, an ideal PBL course is “well organized” but “poorly structured”. The lack of a traditional restrictive structure allows students to pursue and evaluate different solutions to the problem.

The premise for PBL is that actively engaging in problem solving enhances learning in several ways [29, 30]. First, activation of prior knowledge, as occurs in group discussions, aids in learning by providing a framework to incorporate new knowledge. Second, deep processing of material while learning, e.g. by answering questions or using the knowledge, enhances the ability to later recall key concepts. Third, learning in context, e.g. learning the scientific basis for clinical problems in the context of clinical cases, enables and improves recall. These are all effective strategies to enhance learning [31]. PBL opponents argue that acquisition of knowledge is more effective in a traditional didactic curriculum. Further, development of critical thinking skills requires the requisite foundational knowledge to develop realistic solutions to problems [32].

A comprehensive review of PBL outcomes from K-12 through medical school indicated that PBL students perform better in the application of knowledge and reasoning, but not in other areas like basic knowledge [33]. Other recent meta-analyses support the conclusion that PBL, project-based learning and other small group teaching modalities are effective in education from primary school to university, including undergraduate courses in engineering and technology, and pharmacology courses for professional students in health sciences [34–39]. While the majority of the studies reported in these meta-analyses demonstrate that PBL results in better academic performance, there are contrasting studies that demonstrate that PBL is ineffective. This prompts additional investigation to determine the salient factors that distinguish the two outcomes to establish best practices for better results using the PBL platform. Although few studies report the outcomes of PBL based approaches in graduate education, this platform may be beneficial in training biomedical science doctoral students for developing and enhancing critical thinking and practical problem-solving skills.

At our institution, biomedical doctoral students enter an umbrella program and take a core curriculum in the first semester prior to matriculating into one of seven biomedical sciences doctoral programs across a wide range of scientific disciplines in the second semester. Such program diversity created difficulty in achieving consensus on the necessary scientific foundational knowledge for a core curriculum. Common ground was achieved during a recent curriculum revision through the development of required core competencies for all

students, regardless of field of study. These competencies and milestones for biomedical science students at other institutions [5, 14, 15], along with nontraditional approaches to graduate education [12, 25], were used as guidelines for curriculum modification.

Course design

A course was created to develop competencies required by all biomedical sciences doctoral students regardless of their program of interest [14]. As an introductory graduate level course, this met the needs of all our seven diverse biomedical sciences doctoral programs where our first-year doctoral students matriculate. A PBL platform was chosen for the course to engage the students in an active learning environment [25]. The process of problem solving in small teams provided the students with experience in establishing working relationships and how to operate in teams. The students gained experience in researching material from the literature to establish scientific background, find current and appropriate experimental approaches and examples of how results are analyzed. This small group approach allowed each team to develop different hypotheses, experimental plans and analyses based upon the overall interests of the group. The course was designed following discussions with faculty experienced in medical and pharmacy school PBL, and considering course design principles from the literature [27, 40]. The broad learning goals are similar to the overall objectives in another doctoral program using PBL as the primary course format [25], and are aligned with recommended core competencies for PhD scientists [14]. These goals are to:

1. Develop broad, general scientific knowledge (core competency 1 [14]).
2. Develop familiarity with technical approaches specific to each problem.
3. Practice critical thinking/experimental design incorporating rigor and reproducibility,
 - a. including: formulation of hypotheses, detailed experimental design, interpretation of data, statistical analysis (core competencies 3 and 4 [14]).
4. Practice communication skills: written and verbal communication skills (core competency 8 [14]).
5. Develop collaboration and team skills (core competency 6 [14]).
6. Practice using the literature.

Students were organized into groups of four or five based on their scientific background. Student expertise in each group was deliberately mixed to provide different viewpoints during discussion. A single faculty facilitator was assigned to each student group, which met formally in 13 separate sessions (Appendix II). In preparation for each session, the students independently researched topics

using the literature (related to goal 6) and met informally without facilitator oversight to coordinate their findings and organize the discussion for each class session. During the formal one-hour session, one student served as the group leader to manage the discussion. The faculty facilitator guided the discussion to ensure coverage of necessary topics and helped the students identify learning issues, i.e. areas that required additional development, for the students to research and address for the subsequent session. At the end of each session, teams previewed the leading questions for the following class and organized their approach to address these questions prior to the next session. The whole process provided experiences related to goal 5.

As the course was developed during the COVID-19 pandemic, topics related to SARS-CoV2 and COVID-19 were selected as currently relevant problems in society. Session 1 prepared the students to work in teams by discussing about how to work in teams and manage conflict (related to goal 5). In session 2, the students met in their assigned groups to get to know each other, discuss problem-based learning and establish ground rules for the group. Sessions 3 and 4 laid the course background by focusing on the SARS-CoV2 virus and COVID-19-associated pathologies (related to goal 1). The subsequent nine sessions were organized into three separate but interrelated three-session blocks: one on COVID-19 and blood clotting, one on COVID-19 and loss of taste, and one on SARS-CoV2 and therapeutics. The first session in each of these blocks was devoted to covering background information (blood clotting, neurosensation and drug application)(related to goal 1). The second session of each block discussed hypothesis development (mechanisms that SARS-CoV2 infection might utilize to alter blood clotting, the sense of taste, and identification of therapeutic targets to attenuate SARS-CoV2 infection)(related to goal 3). In the second sessions the students also began to design experiments to test the hypothesis. The final session of each block fleshed out the details of the experimental design (related to goals 2 and 3).

The process was iterative, where the students had three opportunities to discuss hypothesis development, experimental design and analysis during sessions with their facilitators. Written and oral presentation assignments (Appendix V) provided additional opportunities to articulate a hypothesis, describe experimental approaches to test the hypotheses, propose analysis of experimental results and develop communication skills (related to goal 4).

Rigor and reproducibility was incorporated into the course. This was an important component given the emphasis recently placed on rigor and reproducibility by federal agencies. As the students built the experimental design to address the hypothesis, recurring

questions were posed to encourage them to consider rigor. Examples include: “*Are the methods and experimental approaches rigorous? How could they be made more rigorous?*” “*Discuss how your controls validate the outcome of the experiment. What additional controls could increase confidence in your result?*” The facilitators were instructed to direct discussion to topics related to the rigor of the experimental design. The students were asked about numbers of replicates, number of animals, additional methods that could be applied to support the experiment, and other measurements to address the hypothesis in a complementary fashion. In the second iteration of the course, we introduced an exercise on rigor and reproducibility for the students using the NIH Rigor and Reproducibility Training Modules (see Appendix III). In this exercise, the students read a short introduction to rigor and reproducibility and viewed a number of short video modules to introduce lessons on rigor. The students were also provided the link to the National Institute of General Medical Sciences clearinghouse of training modules on rigor and reproducibility as reference for experimental design in their future (see Appendix III).

The first delivery of the course was during the COVID-19 pandemic and sessions were conducted on the Zoom platform. The thirteen PBL sessions, and two additional sessions dedicated to oral presentations, were spaced over the course of the first semester of the biomedical sciences doctoral curriculum. The second iteration of the course followed the restructuring of the graduate first year curriculum and the thirteen PBL sessions, plus one additional session devoted to oral presentations, were held during the first three and a half weeks of the first-year curriculum. During this period in the semester, this was the only course commitment for the graduate students. Due to this compressed format, only one written assignment and a single oral presentation were assigned. As the small group format worked well via Zoom in the first iteration of the course, the small groups continued to meet using this virtual platform.

Methods

IRB Approval. The West Virginia University Institutional Review Board approved the study (WVU IRB Protocol#: 2008081739). Informed consent was provided by the participants in writing and all information was collected anonymously.

Surveys. Evaluation of training effectiveness was measured in two ways corresponding to the first two levels of the Kirkpatrick Model of Evaluation [41]. First, students completed a questionnaire upon completion of the course to capture their perceptions of training (Appendix VII). Students were asked their level of agreement/disagreement on a Likert scale with 10 statements about the course and 7 statements about their facilitator. Second,

students took a pre- and post-test to measure differences in their ability to design experiments before and after training (Appendix VIII). The pre- and post-tests were identical, asking the students to design an experiment to test a specific hypothesis, include controls, plan analyses, and state possible results and interpretation. Five questions were provided for the pre- and post-test, where each question posed a hypothesis from a different biomedical discipline, e.g. cancer biology or neuroscience. Students were asked to choose one of the five questions to answer.

Peer-to-peer evaluations were collected to provide feedback on professionalism and teamwork. This survey utilized a Goldilocks scale ranging from 1 to 7, with 4 being the desired score. An example peer question asked about accountability, where responses included *not accountable, e.g. always late* (score=1), *accountable, e.g. punctual, well prepared, follows up* (score=4) and *controlling, e.g. finds fault in others* (score=7). Each student provided a peer-to-peer evaluation for each student in their group. (see Appendix VII). In the second course iteration, Goldilocks surveys were collected three times over the three-week course period due to the compressed time frame. This was necessary to provide rapid feedback to the students about their performance during the course in order to provide opportunities to address and rectify any deficits before making final performance assessments.

Evaluating Pre- and Post-Tests. All pre- and post-test answers were evaluated by three graders in a blind fashion, where the graders were unaware if an answer came from a pre- or post-test. Prior to grading, each grader made up individual answer keys based upon the question(s) on the tests. The graders then met to compare and deliberate these preliminary keys, incorporating changes and edits to produce a single combined key to use for rating answers. While the students were asked to answer one question, some students chose to answer several questions. Superfluous answers were used as a training dataset for the graders. The graders independently scored each answer, then met to review the results and discuss modification of the grading key. The established final grading key, with a perfect score of 16, was utilized by the graders in independently evaluating the complete experimental dataset consisting of all pre- and post-test answers (Appendix IX). To assess the ability of student cohorts to design experiments before and after the course, three of the authors graded all of the pre- and post-test answers. Grading was performed in a blind fashion and the scores of the three raters were averaged for each answer.

Statistical analysis. To measure the interrater reliability of the graders, the intraclass correlation coefficient (ICC) was calculated. A two-way mixed effects model

was utilized to evaluate consistency between multiple raters/measurements. The ICC for grading the training dataset was 0.82, indicating a good inter-rater agreement. The ICC for grading the experimental dataset was also 0.82. For comparison of pre-test vs. post-test performance, the scores of the three raters were averaged for each answer. Since answers were anonymous, the analyses compared responses between individuals. Most, but not all scores, exhibited a Gaussian distribution and therefore a nonparametric statistic, a one-tailed Mann Whitney U test, was used for comparison. The pre-test and post-test scores for 2020 and 2021 could not be compared due to the different format used for the course in each year.

Results

Thirty students participated in the course in the first offering, while 27 students were enrolled in the second year. The students took pre- and post-tests to measure their ability to design an experiment before and after training (Appendix VIII). As the course progressed, students were surveyed on their views of the professionalism of other students in their group (Appendix VII). At the end of the course, students were asked to respond to surveys evaluating the course and their facilitator (see Appendix VII).

Student reception of the course (Kirkpatrick Level 1). In the first year, 23 students responded to the course evaluation (77% response rate) and 26 students submitted facilitator evaluations (87% response rate), whereas in the second year there were 25 responses to the course evaluation (93% response rate) and 26 for facilitators (96% response rate). Likert scores for the 2020 and 2021 course evaluations are presented in Fig. 1. The median score for each question was 4 on a scale of 5 in 2020. In 2021, the median scores for the questions about active learning and hypothesis testing were 5 and the median score of the other questions was 4. The students appreciated the efforts of the facilitators in the course, based upon their evaluations of the facilitators. The median score for every facilitator across all survey questions is shown in Fig. 2. The median score for a single question in 2020 and 2021 was 4.5 and the median score for all other questions was 5. The results of the peer-to-peer evaluations are illustrated in Fig. 3. The average score for each student were plotted, with scores further from the desired score of 4 indicating perceived behaviors that were not ideal. The wide range of scores in the 2020 survey were noted. The students completed three peer-to-peer surveys during the 2021 course. The range of scores in the 2021 peer-to-peer evaluation was narrower than the range in the 2020 survey. The range of scores was expected to narrow from the first (initial) to third (final) survey as students learned and implemented improvements in their

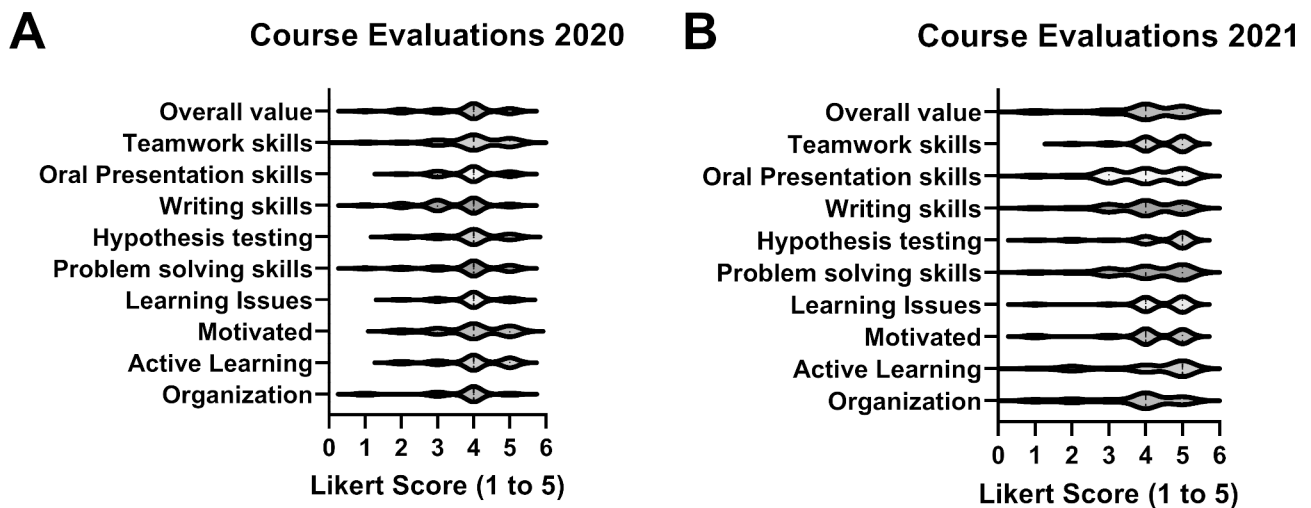


Fig. 1 Results of Course Evaluations by Students. Student evaluations of the course were collected at the end of each offering. The evaluation surveys are in Appendix VII. Violin plots showing the distribution and median score for each question in the 2020 survey (A) and the 2021 survey (B) are shown. The survey used a Likert scale (1 – low to 5 – high)

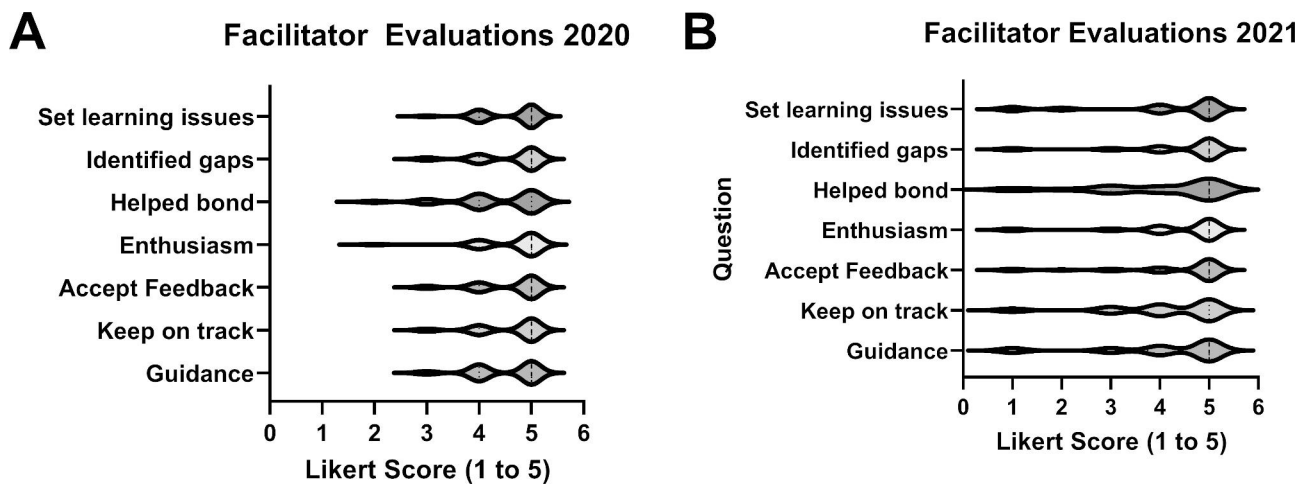


Fig. 2 Results of Facilitator Evaluations by Students. Student evaluations of the facilitators were collected at the end of each offering of the course. The evaluation surveys are in Appendix VII. Violin plots showing the distribution and median score for each question in the 2020 survey (A) and the 2021 survey (B) are shown. The survey used a Likert scale (1 – low to 5 – high)

professional conduct based upon peer feedback. The narrow range of scores in the initial survey left little room for improvement.

Student learning (Kirkpatrick Level 2). Twenty-six students completed the pre-test in each year and consented to participate in this study (87% response in the first year and 96% response in the second year). Eighteen students completed the post-test at the end of the first year (60%) and 26 students completed the test at the end of the second year (96%). Question selection (excluding students that misunderstood the assignment and answered all questions) is shown in Table 1. The most frequently selected questions were Question 1 (45 times) and Question 2 (23 times). Interestingly, the results in Table 1 also indicate that students did not necessarily

choose the same question to answer on the pre-test and post-test.

Average scores on pre-tests and post-tests were compared using a one-tailed Mann Whitney U test. Since the format of the course was different in the two iterations, comparison of test results between the two years could not be made. The average scores of the pre- and post-test in 2020 were not statistically different ($p=0.0673$), although the post-test scores trended higher. In contrast, the difference between the pre- and post-test in 2021 did reach statistical significance ($p=0.0329$). The results collectively indicate an overall improvement in student ability in experimental design (Fig. 4).

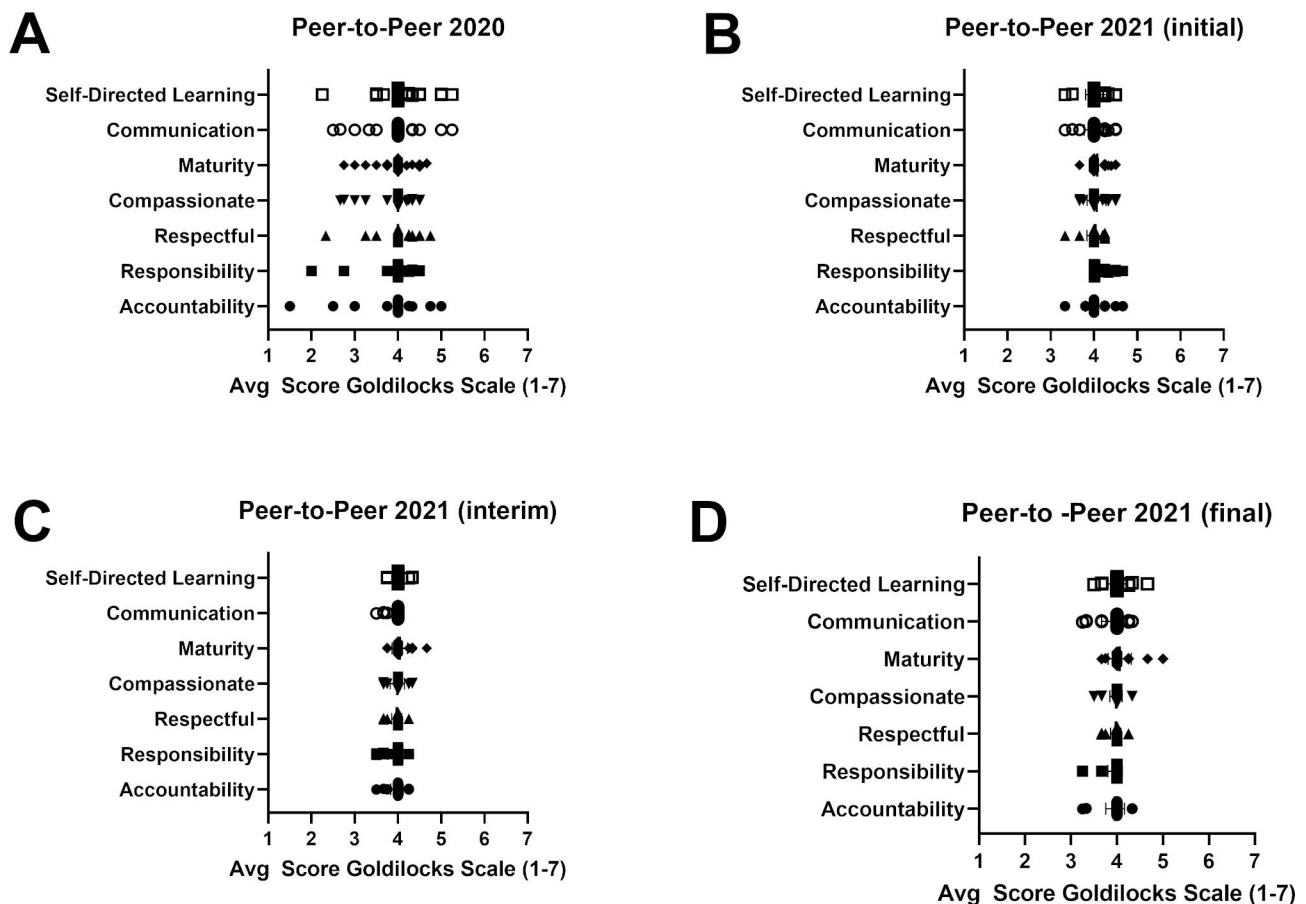


Fig. 3 Results of Student Peer-to-Peer Evaluations. Student peer-to-peer evaluations were collected at the end of the course in year 1 (A), and at the beginning (B), the middle (C) and the end (D) of the course in year 2. Each student evaluated the professionalism of each other student in their group using the evaluation survey shown in Appendix VII. The average score for each student is plotted as a data point. The survey used a Goldilocks scale (range of 1 to 7) where the desired professional behavior is reflected by a score of 4

Table 1 Student Choice of Experimental Question to Answer (Only those who made a choice)

	2020		2021	
	Pre-test	Post-test	Pre-test	Post-test
Body weight	10	8	13	14
Cancer	3	5	7	8
Phosphorylation	2	2	2	3
Vitamin C	4	1	4	0
Neurogenesis	2	0	0	1

Discussion

This course was created in response to biomedical workforce training reports recommending increased training in general professional skills and scientific skills, e.g. critical thinking and experimental design. The course utilizes a PBL format, which is not extensively utilized in graduate education, to incorporate active learning throughout the experience. It was well received by students and analysis suggests that major goals of the course were met. This provides a template for other administrators and

educators seeking to modify curricula in response to calls to modify training programs for doctoral students.

Student evaluations indicated the course was effective at motivating active learning and that students became more active learners. The evaluation survey questions were directly related to three specific course goals: (1) Students reported developing skills in problem solving, hypothesis testing and experimental design. (2) The course helped develop oral presentation skills and written communication skills (in one iteration of the course) and (3) students developed collaboration and team skills. Thus, from the students' perspective, these three course goals were met. Student perceptions of peer professionalism was measured using peer-to-peer surveys. The wide range of Goldilocks scores in the first student cohort was unexpected. In the second student cohort changes in professional behavior were measured over time and the score ranges were narrower. The reasons for the difference between cohorts is unclear. One possibility for this discrepancy is that the first iteration of the course extended over one semester and was during the first

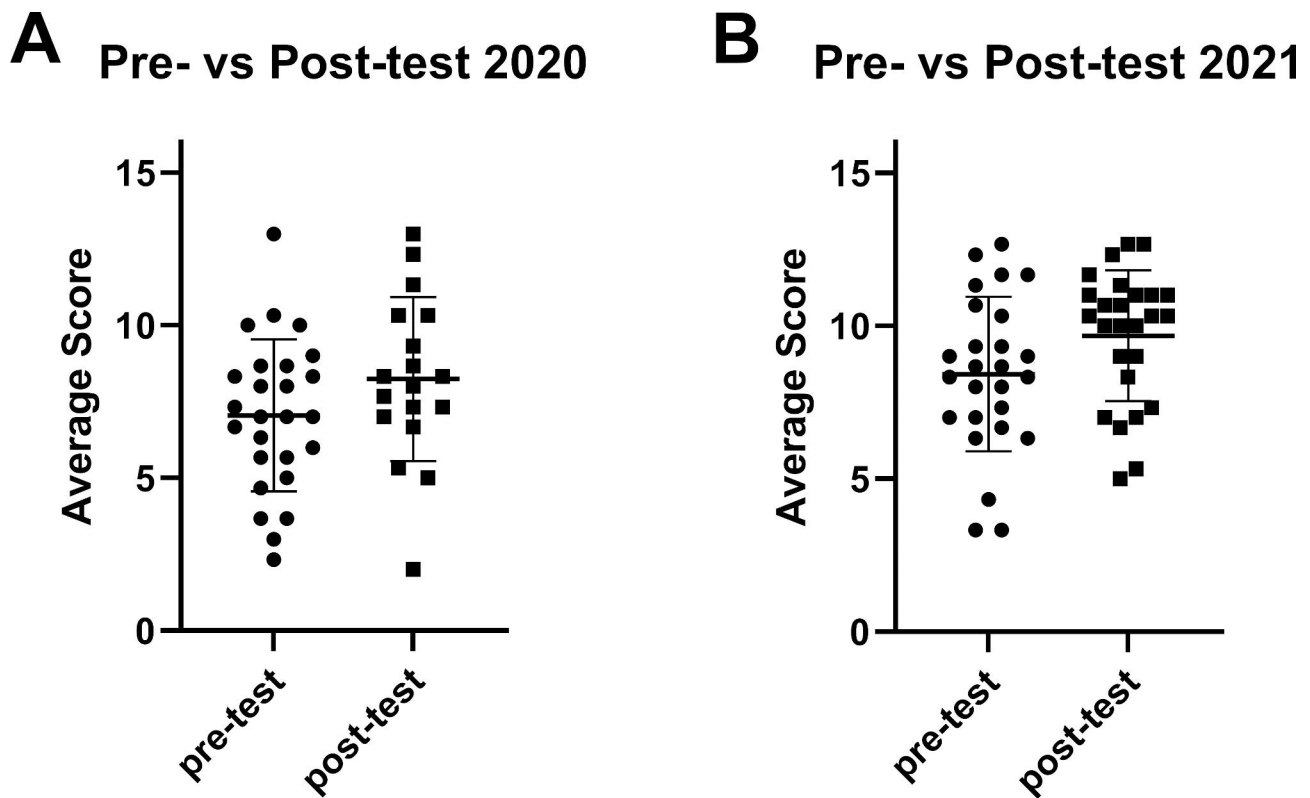


Fig. 4 Pre- and Post-Test Scores. At the beginning and end of each offering, the students completed a test to measure their ability to design an experiment (see Appendix VIII for the details of the exam). Three faculty graded every answer to the pre- and post-test using a common grading rubric (see Appendix IX). The maximum possible score was 16. The average score for each individual answer on the pre-test and post-test is represented as a single data point. The bar indicates the mean score across all answers \pm SD. The average scores of the pre- and post-test scores were compared using a one-tailed Mann Whitney U test. For the 2020 data (A), $p=0.0673$, and for the 2021 data (B), $p=0.0329$

full semester of the pandemic, impacting professional behavior and perceptions of professionalism. The second cohort completed a professionalism survey three times during the course. The narrow range of scores from this cohort in the initial survey made detection of improved professionalism over the course difficult. Results do indicate that professionalism improved in terms of respect and compassion between the first and last surveys. Finally, the results of the pre-test and post-test analysis demonstrated a trend of improved performance on the post-test relative to the pre-test for students in each year of the course and a statistical difference between the pre- and post-test scores in the second year.

Areas for improvement. The course was initially offered as a one-credit course. Student comments on course evaluations and comments in debriefing sessions with facilitators at the end of the course concurred that the work load exceeded that of a one credit course. As a result, the year two version was offered as a two-credit course to better align course credits with workload.

There were student misperceptions about the goals of the course in the first year. Some students equated experimental design with research methods and expressed disappointment that this was not a methods course. While

learning appropriate methods is a goal of the course, the main emphasis is developing hypotheses and designing experiments to test the hypotheses. As such, the choice of methods was driven by the hypotheses and experimental design. This misperception was addressed in the second year by clearly elaborating on the course goals in an introductory class session.

The original course offering contained limited statistical exercises to simulate experimental planning and data analysis, e.g. students were required to conduct a power analysis. Between the first and second years of the course, the entire first semester biomedical sciences curriculum was overhauled with several new course offerings. This new curriculum contained an independent biostatistics workshop that students completed prior to the beginning of this course. Additional statistics exercises were incorporated into the PBL course to provide the students with more experience in the analysis of experimental results. Student evaluations indicated that the introduction of these additional exercises was not effective. Improved coordination between the biostatistics workshop and the PBL course is required to align expectations, better equipping students for the statistical analysis of experimental results encountered later in this course.

An important aspect that was evident from student surveys, facilitator discussions and debrief sessions was that improved coordination between the individual facilitators of the different groups is required to reduce intergroup variability. Due to class size, the students were divided into six groups, with each facilitator assigned to the same group for the duration of the course to maintain continuity. The facilitators met independent of the students throughout the course to discuss upcoming sessions and to share their experiences with their respective groups. This allowed the different facilitators to compare approaches and discuss emerging or perceived concerns/issues. In the second year, one facilitator rotated between different groups during each session to observe how the different student groups functioned. Such a real time faculty peer-evaluation process has the potential to reduce variability between groups, but was challenging to implement within the short three-week time period. Comprehensive training where all facilitators become well versed in PBL strategies and adhere to an established set of guidelines/script for each session is one mechanism that may reduce variability across different facilitator-group pairings.

Limitations. The current study has a number of limitations. The sample size for each class was small, with 30 students enrolled in the first year of the course and 27 students enrolled in the second. The response rates for the pre-tests were high (>87%) but the response rate for the post-test varied between the first year (60%) and second year (96%) of the course. The higher response rate in the second year might be due to fewer end of semester surveys since this was the only course that the students took in that time period. Additionally, the post-test in the second year was conducted at a scheduled time, rather than on the student's own time as was the case in year one. Due to restructuring of the graduate curriculum and the pandemic, the two iterations of the course were formatted differently. This precluded pooling the data from the two offerings and makes comparison between the outcomes difficult.

Presentation of the course was similar, but not identical, to all of the students. Six different PBL groups were required to accommodate the number of matriculating students in each year. Despite efforts to provide a consistent experience, there was variability between the different facilitators in running their respective groups. Further, the development of each session in each group was different, since discussion was driven by the students and their collective interests. These variables could be responsible for increasing the spread of scores on the post-tests and decreasing the value of the course for a subset of students.

The pre- and post-tests were conducted anonymously to encourage student participation. This prevented

correlating the differential between pre- and post-test scores for each student and in comparing learning between different groups. The pre-test and post-test were identical, and provided the students with five options to design experiments (with identical instructions) in response to a different biomedical science problem. An alternative approach could have used isomorphic questions for the pre- and post-tests. It is clear that some students answered the same question on the pre- and post-test, and may benefit from answering the same question twice (albeit after taking the course). Some students clearly answered different questions on the pre- and post-test and the outcomes might be skewed if the two questions challenged the student differently.

While the course analysis captured the first two levels of the Kirkpatrick model of evaluation (reaction and learning), it did not attempt to measure the third level (behavior) or fourth level (results) [41]. Future studies are required to measure the third level. This could be achieved by asking students to elaborate on their experimental design used in recent experiments in their dissertation laboratory following completion of the course, or by evaluating the experimental design students incorporate into their dissertation proposals. The fourth Kirkpatrick level could potentially be assessed by surveying preceptors about their students' abilities in experimental design in a longitudinal manner at semi- or annual committee meetings and accompanying written progress reports. The advantage of focusing on the first two Kirkpatrick levels of evaluation is that the measured outcomes can be confidently attributed to the course. Third and fourth level evaluations are more complicated, since they necessarily take place at some point after completion of the course. Thus, the third and fourth level outcomes can result from additional factors outside of the course (e.g. other coursework, working in the lab, attendance in student-based research forum, meeting with mentors, etc.). Another limiting factor is the use of a single test to measure student learning. Additional alternative approaches to measure learning might better capture differences between the pre- and post-test scores.

Implementation. This curriculum is readily scalable and can be modified for graduate programs of any size, with the caveat that larger programs will require more facilitators. At Van Andel, the doctoral cohorts are three to five new students per year and all are accommodated in one PBL group [25]. At our institution, we have scaled up to a moderate sized doctoral program with 25 to 30 matriculating students per year, dividing the students into six PBL groups (4–5 students each). Medical School classes frequently exceed 100 students (our program has 115–120 new students each fall) and typically have between five and eight students per group. Our graduate course has groups at the lower end of this range. This

course could be scaled up by increasing the number of students in the group or by increasing the number of groups.

Consistency between groups is important so each group of students has a similar experience and reaps the full benefit of this experience. Regular meetings between the course coordinator and facilitators to discuss the content of upcoming sessions and define rubrics to guide student feedback and evaluation were mechanisms used to standardize between the different groups in this course (Appendix VI). In hindsight, the course would benefit from more rigorous facilitator training prior to participation in the course. While a number of our facilitators were veterans of a medical school PBL course, the necessary skillset required to effectively manage a graduate level PBL course that is centered on developing critical thinking and experimental design are different. Such training requires an extensive time commitment by the course coordinators and participating facilitators.

The most difficult task in developing this course involved the course conception and development of the problem-based assignments. Designing a COVID-19 based PBL course in 2020 required *de novo* development of all course material. This entailed collecting and compiling information about the virus and the disease to provide quick reference for facilitators to guide discussion in their groups, all in the face of constantly shifting scientific and medical knowledge, along with the complete lack of traditional peer-based academic social engagement due to the pandemic. In development of this course, three different COVID-based problems were identified, with appropriate general background material for each problem requiring extensive research and development. Background material on cell and animal models, general strategies for experimental manipulation and methods to measure specific outcomes were collected in each case. Student copies for each session were designed to contain a series of questions as a guide to identifying important background concepts. Facilitator copies for each session were prepared with the goal of efficiently and effectively guiding each class meeting. These guidelines contained ideas for discussion points, areas of elaboration and a truncated key of necessary information to guide the group (Appendix IV). Several PBL repositories exist (e.g. <https://itue.udel.edu/pbl/problems/>, <https://www.nsta.org/case-studies>) and MedEdPORTAL (<https://www.mededportal.org/>) publishes medical-specific cases. These provide valuable resources for case-based ideas, but few are specifically geared for research-focused biomedical graduate students. As such, modification of cases germane to first year biomedical graduate students with a research-centered focus is required prior to implementation. Finally, appropriate support materials for surveys and evaluation rubrics requires additional development

and refinement of current or existing templates to permit improved evaluation of learning outcomes (Appendix VI).

Development of an effective PBL course takes considerable time and effort to conceive and construct. Successful implementation requires the requisite higher administrative support to identify and devote the necessary and appropriate faculty needed for course creation, the assignment of skilled faculty to serve as facilitators and staff support to coordinate the logistics for the course. It is critical that there is strong faculty commitment amongst the facilitators to devote the time and energy necessary to prepare and to successfully facilitate a group of students. Strong institutional support is linked to facilitator satisfaction and commitment to the PBL-based programs [42]. Institutional support can be demonstrated in multiple ways. The time commitment for course developers, coordinators and facilitators should be accurately reflected in teaching assignments. Performance in these roles in PBL should factor into decisions about support for professional development, e.g. travel awards, and merit based pay increases. Further, efforts in developing, implementing and executing a successful PBL course should be recognized as important activities during annual faculty evaluations by departmental chairs and promotion and tenure committees.

Key Takeaways. The creation and implementation of this course was intellectually stimulating and facilitators found their interactions with students gratifying. From student survey responses and test results the course was at least modestly successful at achieving its goals. Based upon our experience, important issues to consider when deciding to implement such a curriculum include: (1) support of the administration for developing the curriculum, (2) facilitator buy-in to the approach, (3) continuity (not uniformity) between PBL groups, (4) other components of the curriculum and how they might be leveraged to enhance the effectiveness of PBL and (5) effort required to develop and deliver the course, which must be recognized by the administration.

Future Directions. Novel curriculum development is an often overlooked but important component to contemporary graduate student education in the biomedical sciences. It is critical that modifications incorporated in graduate education are evidence based. We report the implementation of a novel PBL course for training in the scientific skill sets required for developing and testing hypotheses, and demonstrate its effectiveness. Additional measures to assess the course goals in improving critical thinking, experimental design and self-efficacy in experimental design will be implemented using validated tests [22, 43–45]. Further studies are also required to determine the long-term impact of this training on student performance in the laboratory and progression towards

degree. It will be interesting to determine if similar curriculum changes to emphasize development of skills will shorten the time to degree, a frequent recommendation for training the modern biomedical workforce [1, 46–48].

Incorporation of courses emphasizing development of skills can be done in conjunction with traditional didactic instruction to build the necessary knowledge base for modern biomedical research. Our PBL course was stand-alone, requiring the students to research background material prior to hypothesis development and experimental design. Coordination between the two modalities would obviate the need for background research in the PBL component, reinforce the basic knowledge presented didactically through application, and prepare students for higher order thinking about the application of the concepts learned in the traditional classroom. Maintaining a balance between problem-based and traditional instruction may also be key in improving faculty engagement into such new and future initiatives. Continued investments in the creation and improvement of innovative components of graduate curricula centered around developing scientific skills of doctoral students can be intellectually stimulating for faculty and provide a better training environment for students. The effort may be rewarded by streamlining training and strengthening the biomedical workforce of the future.

Abbreviations

PBL	Problem-based learning
STEM	Science, technology, engineering, and math
K-12	kindergarten through grade 12
ICC	Intraclass coefficient >
SARS-CoV2	severe acute respiratory syndrome coronavirus 2
COVID-19	Coronavirus disease 19

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12909-023-04569-7>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5

Acknowledgements

Thanks to Mary Wimmer and Drew Shiemke for many discussions over the years about PBL in the medical curriculum and examples of case studies. We thank Steve Treisenberg for initial suggestions and discussions regarding PBL effectiveness in the Van Andel Institute. Thanks to Paul and Julie Lockman for discussions about PBL in School of Pharmacy curricula and examples of case studies. Special thanks to the facilitators of the groups, Stan Hileman, Hunter Zhang, Paul Chantler, Yehenev Agazie, Saravan Kolandaivelu, Hangang Yu, Tim Eubank, William Walker, and Amanda Gatesman-Ammer. Without their considerable efforts the course could never have been successfully implemented. Thanks to the Department of Biochemistry and Molecular Medicine for supporting the development of this project. MS is the director of the Cell & Molecular Biology and Biomedical Engineering Training Program (T32 GM133369).

Author contributions

SW and MS developed the concept for the course. MS was responsible for creation and development of all of the content, for the implementation of the course, the design of the study and creating the first draft of the manuscript. MG, MRG and SW graded the pre- and post-test answers in a blind fashion. MS, MG, MRG and SW analyzed the data and edited the manuscript.

Funding

There was no funding available for this work.

Data Availability

All data generated in this study are included in this published article and its supplementary information files.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

The West Virginia University Institutional Review Board approved the study (WVU IRB Protocol#: 2008081739). Informed consent was provided in writing and all information was collected anonymously. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable.

Received: 4 March 2023 / Accepted: 5 August 2023

Published online: 16 August 2023

References

- National Institutes of Health. Biomedical research workforce working group report. Bethesda, MD: National Institutes of Health; 2012.
- Sinche M, Layton RL, Brandt PD, O'Connell AB, Hall JD, Freeman AM, Harrell JR, Cook JG, Brennwald PJ. An evidence-based evaluation of transferrable skills and job satisfaction for science PhDs. *PLoS ONE*. 2017;12:e0185023.
- Ghaffarzadegan N, Hawley J, Larson R, Xue Y. A note on PhD Population Growth in Biomedical Sciences. *Syst Res Behav Sci*. 2015;23:402–5.
- National Academies of Sciences Engineering and Medicine. The next generation of biomedical and behavioral sciences researchers: breaking through. Washington, DC: National Academies Press (US); 2018.
- National Academies of Sciences Engineering and Medicine. Graduate STEM education for the 21st century. Washington, DC: National Academies Press; 2018.
- Roach M, Sauermaun H. The declining interest in an academic career. *PLoS ONE*. 2017;12:e0184130.
- Sauermaun H, Roach M. Science PhD career preferences: levels, changes, and advisor encouragement. *PLoS ONE*. 2012;7:e36307.
- St Clair R, Hutto T, MacBeth C, Newstetter W, McCarty NA, Melkers J. The "new normal": adapting doctoral trainee career preparation for broad career paths in science. *PLoS ONE*. 2017;12:e0177035.
- Fuhrmann CN, Halme DG, O'Sullivan PS, Lindstaedt B. Improving graduate education to support a branching career pipeline: recommendations based on a survey of doctoral students in the basic biomedical sciences. *CBE—Life Sci Educ*. 2011;10:239–49.
- Casadevall A, Ellis LM, Davies EW, McFall-Ngai M, Fang FC. (2016) A framework for improving the quality of research in the biological sciences. 7, e01256–01216.
- Casadevall A, Fang FC. (2016) Rigorous science: a how-to guide. 7, e01902–01916.
- Bosch G, Casadevall A. Graduate Biomedical Science Education needs a New Philosophy. *mBio*. 2017;8:e01539–01517.
- Bosch G. Train PhD students to be thinkers not just specialists. *Nature*. 2018;554:277–8.
- Verderame MF, Freedman VH, Kozlowski LM, McCormack WT. (2018) Competency-based assessment for the training of PhD students and early-career scientists. *Elife* 7, e34801.

15. Graziane J, Graziane N. Neuroscience Milestones: developing standardized core-competencies for Research-Based neuroscience trainees. *J Neurosci*. 2022;42:7332–8.
16. Edgar L, Roberts S, Holmboe E. Milestones 2.0: a step forward. *J graduate Med Educ*. 2018;10:367–9.
17. Kiley M, Wisker G. Threshold concepts in research education and evidence of threshold crossing. *High Educ Res Dev*. 2009;28:431–41.
18. Timmerman BC, Feldon D, Maher M, Strickland D, Gilmore J. Performance-based assessment of graduate student research skills: timing, trajectory, and potential thresholds. *Stud High Educ*. 2013;38:693–710.
19. Lachance K, Heustis RJ, Loparo JJ, Venkatesh MJ. Self-efficacy and performance of research skills among first-semester bioscience doctoral students. *CBE—Life Sci Educ*. 2020;19:ar28.
20. Heustis RJ, Venkatesh MJ, Gutlierner JL, Loparo JJ. Embedding academic and professional skills training with experimental-design chalk talks. *Nat Biotechnol*. 2019;37:1523–7.
21. Ulibarri N, Cravens AE, Cornelius M, Royalty A, Nabergoj AS. Research as design: developing creative confidence in doctoral students through design thinking. *Int J Doctoral Stud*. 2014;9:249–70.
22. Gottesman AJ, Hoskins SG. CREATE cornerstone: introduction to scientific thinking, a new course for STEM-interested freshmen, demystifies scientific thinking through analysis of scientific literature. *CBE—Life Sci Educ*. 2013;12:59–72.
23. Koenig K, Schen M, Edwards M, Bao L. (2012) Addressing STEM Retention through a scientific thought and methods Course. *J Coll Sci Teach* 41.
24. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP. Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci*. 2014;111:8410–5.
25. Turner JD, Triesenberg SJ. PBL for Ph. D.: a problem-based learning approach to doctoral education in biomedical research. *ASQ High Educ Brief*. 2010;3:1–5.
26. Neufeld VR, Barrows HS. The “McMaster Philosophy”: an approach to medical education. *Acad Med*. 1974;49:1040–50.
27. Duch BJ, Groh SE, Allen DE. The power of problem-based learning: a practical how to for teaching undergraduate courses in any discipline. Sterling, VA: Stylus Publishing, LLC; 2001.
28. Wirkala C, Kuhn D. Problem-based learning in K–12 education: is it effective and how does it achieve its effects? *Am Educ Res J*. 2011;48:1157–86.
29. Norman G, Schmidt HG. The psychological basis of problem-based learning: a review of the evidence. *Acad Med*. 1992;67:557–65.
30. Handelsman J, Ebert-May D, Beichner R, Bruns P, Chang A, DeHaan R, Gentile J, Lauffer S, Stewart J, Tilghman SM. Scientific teaching. *Science*. 2004;304:521–2.
31. Brown PC, Roediger III, H. L., and, McDaniel MA. Make it stick: the science of successful learning. Cambridge, Massachusetts: The Belknap Press of Harvard University Press; 2014.
32. Willingham DT. Critical thinking: why is it so hard to teach? *Arts Educ Policy Rev*. 2008;109:21–32.
33. Hung W, Jonassen DH, Liu R. (2008) Problem-based learning. In *Handbook of research on educational communications and technology* pp. 485–506, Routledge, Abingdon UK.
34. Uluçınar U. The Effect of Problem-Based learning in Science Education on Academic Achievement: a Meta-Analytical Study. *Sci Educ Int*. 2023;34:72–85.
35. Chen C-H, Yang Y-C. Revisiting the effects of project-based learning on students’ academic achievement: a meta-analysis investigating moderators. *Educational Res Rev*. 2019;26:71–81.
36. Liu Y, Pásztor A. Effects of problem-based learning instructional intervention on critical thinking in higher education: a meta-analysis. *Think Skills Creativity*. 2022;45:101069.
37. Kalaian SA, Kasim RM, Nims JK. Effectiveness of small-group learning pedagogies in Engineering and Technology Education: a Meta-analysis. *J Technol Educ*. 2018;29:20–35.
38. Liu L, Du X, Zhang Z, Zhou J. Effect of problem-based learning in pharmacology education: a meta-analysis. *Stud Educational Evaluation*. 2019;60:43–58.
39. Dochy F, Segers M, Van den Bossche P, Gijbels D. Effects of problem-based learning: a meta-analysis. *Learn instruction*. 2003;13:533–68.
40. Azer SA. Challenges facing PBL tutors: 12 tips for successful group facilitation. *Med Teach*. 2005;27:676–81.
41. Kirkpatrick DL. Seven keys to unlock the four levels of evaluation. *Perform Improv*. 2006;45:5–8.
42. Trullàs JC, Blay C, Sarri E, Pujol R. Effectiveness of problem-based learning methodology in undergraduate medical education: a scoping review. *BMC Med Educ*. 2022;22:104.
43. Deane T, Nomme K, Jeffery E, Pollock C, Birol G. Development of the biological experimental design concept inventory (BEDCI). *CBE—Life Sci Educ*. 2014;13:540–51.
44. Sirum K, Humburg J. The experimental design ability test (EDAT). *Bioscene: J Coll Biology Teach*. 2011;37:8–16.
45. Hoskins SG, Lopatto D, Stevens LM. The CREATE approach to primary literature shifts undergraduates’ self-assessed ability to read and analyze journal articles, attitudes about science, and epistemological beliefs. *CBE—Life Sci Educ*. 2011;10:368–78.
46. Pickett CL, Corb BW, Matthews CR, Sundquist WI, Berg JM. Toward a sustainable biomedical research enterprise: finding consensus and implementing recommendations. *Proc Natl Acad Sci U S A*. 2015;112:10832–6.
47. National Research Council. Research universities and the future of America: ten breakthrough actions vital to our nation’s prosperity and security. Washington, DC: National Academies Press; 2012.
48. American Academy of Arts and Sciences. Restoring the Foundation: the vital role of Research in preserving the American Dream: report brief. American Academy of Arts & Sciences; 2014.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.