Educational Approach to Cyber Foundations in an Undergraduate Core Program

Jason P. Hussey  
*United States Military Academy, jason.hussey@usma.edu*

Jacob E. Shaha  
*United States Military Academy, jacob.shaha@usma.edu*

Follow this and additional works at: [https://digitalcommons.usmalibrary.org/eecs_rp](https://digitalcommons.usmalibrary.org/eecs_rp)

🔗 Part of the [Databases and Information Systems Commons](https://digitalcommons.usmalibrary.org/eecs_rp), [Educational Methods Commons](https://digitalcommons.usmalibrary.org/eecs_rp), and the [Science and Mathematics Education Commons](https://digitalcommons.usmalibrary.org/eecs_rp)

**Recommended Citation**

Educational Approach to Cyber Foundations in an Undergraduate Core Program

Jason Hussey and Jacob Shaha
United States Military Academy
West Point, NY 10996
USA
jason.hussey@usma.edu, jacob.shaha@usma.edu

ABSTRACT
Maximizing a student’s learning in a general education information technology course is critical when teachers have only a little time to cover numerous topics within the discipline. It is therefore paramount that programs utilize the most effective pedagogical approach to educating students on these topics. This allows teachers to take full advantage of this limited time per topic. The principal contribution of this paper is a statistical analysis of student performance in an intermediate-level information technology course when exposed to two popular methods of teaching information technology concepts. This course is part of the larger cyber education model at the United States Military Academy. Our study implemented and analyzed the results from a control group educated with systematic, skills-based instruction versus a treatment group where a problem-centered learning approach was utilized. Our experimental results provide statistically significant reinforcement of the idea that problem-centered learning is superior to skills-based instruction for educating students in information technology topics as a part of their cyber education.

KEYWORDS
ACM proceedings; cyber; cyber education; problem-centered learning; multi-discipline cyber education; general education requirement; core program; database education; databases

1 INTRODUCTION
When information technologies become a mainstay of any discipline, undergraduate programs that prepare future professionals in those disciplines must develop curriculum educating students on the concepts and application of these technologies. Now, with the ubiquity of data across virtually every industry and the emergence of the cognitive computing era, it is important that all undergraduates emerge from their educational experience armed with an understanding of cyberspace and the tools within to achieve one’s organizational goals.

Covering the main topics of a given discipline in a single semester is a major challenge for any type of general education or core course. We often refer to these courses as ‘survey’ courses, highlighting their emphasis on breadth, not depth. The necessarily limited time that instructors get to spend on any one topic can lead to overly brief exposure for the students and thus little knowledge acquired in the subject.

Adding to the challenges of teaching information technology to a large portion of all undergraduates, Christopher Brown et al. point out that students often enter the course with a lower level of base computer skills than expected [1]. When instructors have to spend time building these base computer skills, it further detracts from time spent with the subject matter.

It is clear that a cyber component is critical to every modern undergraduate’s education. What is not always clear is the most effective way to utilize the limited time in each module to maximize student learning. Therefore, we demonstrate how different pedagogical approaches to information technology topics in a general education course can maximize the performance of students given the limited time we have in any individual module. Our research compares a problem-centered learning approach with the more common systematic, skills-based method for learning information technology concepts and applications. Specifically, we present a statistical analysis of 376 students’ performance in the database module of an intermediate-level general education course in Cyber Foundations, based on whether they received problem-centered or skills-based instruction.

2 RELATED WORK
Throughout the first decade of the 2000s, Dr. Catherine Chen authored several papers concerning pedagogical approaches to teaching databases to business students. In 2004, Chen and Ray identified elements of systematic, step-by-step instruction that hold students back from being able to apply their knowledge to new situations [2]. The subjects of her study “stated that it was easy to follow the instructions in the book” but had trouble applying the same techniques later on their own [2]. She also concluded that for students to recognize the capabilities that databases provide for decision-making and problem-solving, it is vital that there be real-world context to the problems they are
asked to solve in class. Our own study builds upon this work in that we seek to measure the actual improvement in student performance and confidence when exposed to problem-centered learning as opposed to skills-based instruction.

In 2010, Chen revisited the topic of teaching database skills and problem-solving with databases [3]. She studied the impact that different pedagogical approaches had on problem-solving transfer capabilities in business students. She concludes that problem-solving instruction and problem-solving discovery had no impact on a student’s ability to transfer their knowledge to a new problem. Chen suggests that there are three levels in which students need proficiency in order to be able to transfer their skills to new problems. The first is learning simple procedures required to manipulate databases. The second is knowing when and why to apply these procedures. The final level is understanding problem-solving processes. When considering the time constraints in a survey-style course, it’s possible students will not be able to progress through the first two tiers that Chen suggests are necessary for problem-solving transfer.

Chen points out that more research is needed as her conclusion runs counter to what Phye suggests in a similar study in 2001. In that study, Phye demonstrates and concludes that “without doubt, instruction and practice with problem-solving strategies promote problem-solving transfer” [4].

On the other hand, some studies suggest that for an introductory course teaching highly structured topics where students have little-to-no prior knowledge of the subject matter, worked examples are superior to problem-solving instruction. Halabi and Tuovinen studied the effects of these two pedagogical approaches on students participating in computer-based learning for an introductory accounting course [5]. They conclude that the reduced cognitive load required for the worked examples led to higher instructional efficiency. This approach is supported by Kalyuga et al. when learners have no prior knowledge of the subject [6].

Acknowledging that there is some precedent for implementing problem-based or problem-centered learning in information technology curricula, several papers have gone on to explore what that looks like in the classroom. Laware and Walters provide the cyber education community with a thorough review of lessons learned in their course over a three year period where they experimented with various active learning strategies [7].

L’Heureux et al. describe a course in which they implemented information technology problem solving to develop computational thinking in students [8]. Where many of these types of studies evaluated the suitability of this problem-centered approach via student feedback on surveys, we seek to measure it against more traditional methods of teaching information technology via statistical analysis of data from course assessments.

Christopher Brown et al. provide some helpful lessons learned from the process of constructing and delivering a cybersecurity curriculum in a core sequence at the United States Naval Academy. They conclude that a technically oriented core course would benefit from more “real-world contextual reinforcement of the technical concepts discussed” [1]. Whether or not introducing real-world problems into the curriculum will tangibly improve student performance is largely what we demonstrate in this paper.

In this study, we build upon these and other works by analyzing student performance when exposed to problem-centered learning pedagogy. This study was implemented in a core requirements course with 376 student samples. Using this large sample set, we conduct statistical analysis of student academic performance in order to assess whether problem-centered learning has a significant, measurable impact as compared to more structured, skills-based instruction typical of information technology courses.

3 BACKGROUND

The ideas and focus of this paper are nested within the multi-level, multi-discipline approach to cyber education presented by Sobiesk et al. [9]. The authors set forth a model for cybersecurity education in an undergraduate curriculum. This model is based on the implementation at the United States Military Academy (USMA), the same institution at which we teach. Sobiesk describes integrating aspects of cyber education across an undergraduate student’s entire educational experience. The goal of such an approach is to ensure that graduates have the "foundational knowledge, skills, and abilities needed to succeed in the 21st Century Cyber Domain” [9].

Our work presents methods successfully used within this paradigm. We focus on the base level of the cyber education model depicted in Figure 1, the general education program. Specifically, this study centered on an intermediate-level information technology course required of approximately two-thirds of the junior class at USMA. This course is the second of a two-course sequence that comprises the general education component of cyber education at USMA.

4 METHODOLOGY: COURSE STRUCTURE AND PEDAGOGY

We hypothesized that a problem-centered learning approach to teaching databases to non-STEM majors (hereafter, “non-majors”) in a constrained timeframe would result in better understanding of how and when to use databases to meet organizational goals. Further, we hypothesized that students in
the treatment group would have a higher degree of self-reported confidence in recognizing situations where databases could help them solve organizational problems.

4.1 The Course

The course used for this study, CY305 - Cyber Foundations, is a core requirement for juniors at USMA. CY305 is the second of a two-course sequence required of most non-majors at USMA. Therefore, students in CY305 have already taken the introductory-level information technology course, IT105, where the emphasis is on introductory computer programming. IT105 also briefly exposes students to computer architecture, networks, and security and privacy.

CY305 is primarily populated by non-majors. Those that do not take our course are generally in technically-oriented majors and satisfy the cyber education goals of the institution with another, more focused course. Enrollment is between 350 and 400 students during the fall and spring semesters. The course educates students on leveraging information technology to achieve organizational goals and improve decision-making. Topics in this course include digitization, web development, databases, computer networks, and concepts in cyber security.

For this study, we focused on the database module of the course. This module is comprised of six 55-minute lessons and a single 55-minute project work period. Students gain exposure to the concept of relational databases followed by hands-on experience working with Microsoft Access, a desktop database management system.

4.2 The Control Group

In its current format, the unit takes a systematic, skills-based approach to teaching students how to use Microsoft Access. Each lesson, the teacher discusses a set of concepts and demonstrates corresponding features in the application at the beginning of the lesson. Following the demonstration, the students conduct a practical exercise in class. These exercises consist of step-by-step instructions on how to accomplish various tasks within the program and are designed to expose students to the concepts and features presented by the instructor.

4.3 The Treatment Group

We designed a problem-centered curriculum around three problem scenarios. One of the established goals of this curriculum revision was to maximize students’ exposure to different situations where database solutions could be applied. Each of the three scenarios were explored in various portions of the unit: one scenario was used for in-class lectures and demonstrations, another was used for homework assignments, and the third was used for the database project.

The in-class scenario is modeled after cadet life at USMA to lend immediate familiarity with the problems being described. For homework assignments, students were given a commercial business scenario to familiarize them with transferring concepts and skills learned during in-class exercises to a new and slightly different situation. Finally, the culminating project introduces a third problem environment modeled after their future careers as junior officers in the U.S. Army.

At the start of each lesson in the treatment groups, instructors spent 10-15 minutes posing questions to students to illuminate the problem that the class was to solve that day. As an example, in the first lesson students received several spreadsheets containing related data and were asked to answer a series of increasingly complex questions, requiring information from multiple spreadsheets. As the instructor asked more complicated questions, the students realized that the act of retrieving information from many discrete spreadsheets can quickly become a challenge. Armed with that realization and understanding the nature of the problem, the students were then introduced to the basic concepts of relational databases. Then, the students imported the data into a database and constructed the corresponding relationships. After a quick primer on query building, the students were able to begin producing answers to the instructor’s earlier, complex questions. They recognized the power of the tool and its potential for application in their own careers as students and, eventually, as military officers.

4.4 The Assessments

Student learning and understanding was assessed through performance on projects and tests covering the database module’s material. In order to gauge the impact on the performance of the treatment group as compared to that of the control group, we held all of the major graded assessments constant between the two. The primary change for the treatment group was the pedagogical approach in the classroom. Additionally, it was necessary to modify the content of the in-class and homework assignments to better fit the pacing and ordering of topics under the treatment curriculum.

4.4.1 Short-term assessments. Immediately following the completion of the database module of CY305, students complete an approximately 4-6 hour project. This project requires students to answer specific questions by designing queries against a starting database. Students then expand the database given a design diagram. Finally, they are required to write more complex queries against the newly expanded database.

This project is designed to test students’ ability to independently apply what they have learned in the database module to a real-world scenario modeled around their future careers as military officers.

In addition to a hands-on project, students are evaluated on their understanding of database concepts through 20 multiple-choice style questions on a mid-term examination. These questions specifically focus on the theory behind databases and database design, leaving the practical application for the project.

4.4.2 Long-term assessments. As part of the 3.5-hour final examination, students complete another set of multiple-choice questions. One-sixth of these questions cover high-level database concepts. This representation of database questions corresponds to the percentage of CY305’s overall curriculum that databases occupied.
As the next part of their final examination, students have to complete a much simpler version of the database practical application. Tasks include expanding a small database and developing two moderately difficult queries to answer specific questions.

5 EXPERIMENTAL DESIGN

5.1 Sample Selection

CY305 is taught to multiple sections of 16-18 students, by multiple instructors, using a single unified curriculum. Lesson presentation varies per instructor, but all assignments, exams, and projects are standardized between sections.

At the outset of each semester, CY305 students are grouped by Cumulative Quality Point Average (CQPA), an academic performance measure analogous to Grade Point Average (GPA). Students are then randomly assigned to sections such that like-CQPA students are tracked together; the assignment algorithm minimizes the CQPA variance within a single section. The result is sections of pseudo-randomly assigned students with highly similar academic performance records. These sections are the experimental units in this study.

Instructors are then assigned to sections based on schedule availability and previous experience with the curriculum; more experienced instructors are generally assigned to sections with lower mean CQPAs.

The cohort sections for our experimental groups were generated in this manner, at the beginning of the semester. Prior to commencing the database module of the course, we designated instructors to teach either the control curriculum or the treatment curriculum. To prevent crossover effects between the treatment and control curricula, each instructor was assigned to only one of the two versions, to be used in all that instructor’s sections.

Instructors were assigned to satisfy three criteria:

1. The student sample size between the treatment and control groups should be roughly equivalent. The resulting sample sizes were: $N_{\text{ctrl}} = 207, N_{\text{mnt}} = 169$.
2. The treatment and control groups should have roughly equivalent numbers of high-performing sections and low-performing sections. The control group and treatment group included both high-CQPA and low-CQPA sections, with high defined as a 3.0 or above CQPA. However, the previously stated constraint (a given instructor teaches only one version of the curriculum) prevented full balance in this regard. More low-performing sections were placed in the treatment group than the control group, and as a result, the control group had a higher mean CQPA than the treatment group, as summarized in Table 1.
3. The treatment groups were specifically assigned to instructors with a strong knowledge of both the CY305 curriculum and the general theory and application of databases. Conversely, the most experienced instructors – those with the most previous iterations of CY305 – were deliberately placed in the control group. This was done to ensure that the treatment curriculum was presented as strongly as the control curriculum.

<table>
<thead>
<tr>
<th>Table 1: CQPA Descriptive Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cohort</strong></td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>Treatment</td>
</tr>
</tbody>
</table>

5.2 Outcomes

We measured two primary outcomes: raw academic scores, and self-reported confidence and comfort scores.

5.2.1 Academic Scores. Academic scores were collected on two occasions, via two separate methods. Short-term scores were assessed immediately following the conclusion of the teaching block, via a multiple-choice exam and a database project. The exam and project together assessed a student’s familiarity with database principles, and with the specifics of Microsoft Access database implementation. Long-term scores were assessed using a similar multiple-choice exam and a simplified database practical application, of the same type as the initial project but of greatly reduced scope. These long-term assessments were conducted roughly six weeks after the conclusion of the instruction block, during final examinations.

In the analysis, short-term, long-term, and total raw scores were considered only in aggregate; scores from a respective interval’s exam and practical application were summed into a single cumulative score for that interval. We refer to this total point value as the “raw” score. Raw academic scores are represented as a percentage, total points earned over total points possible.

5.2.2 Grade Ratio. As an additional academic outcome, we calculated each student’s “grade ratio.” This was calculated by translating the student’s raw numeric score to the corresponding grade value on a 4.33-point scale. This “effective grade” was then divided by the student’s CQPA to find the student’s grade ratio. The CQPA represents the average of the student’s previous performance and thus can be interpreted as a predictor of the student’s CY305 performance. To ensure the validity of this approach, we calculated the Pearson’s correlation coefficient between CQPA and the student’s total effective grade (considering all exams and projects). The calculated value was $\rho=0.46, p<0.000$. The CQPA is strongly correlated to the student’s final grade and therefore serves as a strong estimate of a student’s performance in the course.

The grade ratio, as calculated, is the proportion of the student’s anticipated performance that was actually achieved. So, a student with a 3.00 CQPA (a B average) who achieved an 85% (a B) in the raw academic scores would have a grade ratio of 1.0, having performed as predicted by the CQPA. A grade ratio of greater than 1.0 indicates the student performed better than their historical performance would have predicted. Conversely, a student with less than a 1.0-grade ratio failed to perform as well as expected.

The use of this calculated variable was a deliberate a priori decision to compensate for the disparity in CQPA between the cohorts. This metric better captures how much improvement was seen in each student, relative to that student’s previously measured capability. We hypothesized that raw scores would be equivalent between cohorts – indeed that was a design specification, as it would be unethical to disadvantage either cohort in an academic
setting. However, achieving raw numeric equivalence would be a greater accomplishment if one cohort was lower-performing on average. As this was the case, the grade ratio was devised as an effective measurement of the treatment’s academic impact.

5.2.3 Confidence Scores. Confidence and comfort scores measured a student’s self-reported familiarity, confidence, and ease with databases as an information technology tool. Scores were collected using an online survey. Hypothetical scenarios requiring complex information tracking and reporting were presented. Students then used the Likert scale to report their confidence and preparedness to succeed in those scenarios. These reported scores were used to calculate each student’s “composite confidence”, the normalized average of each Likert response. We designed the survey such that each question was asking for a similar measurement using different wording, with the intent of averaging the resulting scores to produce a general metric for a student’s self-assessed capability with databases.

6 RESULTS AND ANALYSIS
All statistical analyses were done using Minitab 17.2.1 [10].

6.1 Raw Academic Scores
Raw academic scores (denoted “Raw” in Table 2 and described in 5.2.1) were compared between cohorts using two-sample T-tests. Raw scores are computed as the percentage of points earned on the assessments for the specified interval. Comparisons were made for raw scores in the short-term, long-term, and the total of the two scores. The results are summarized in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>Short-term</th>
<th>Long-term</th>
<th>Course Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( GR_{CTRL} )</td>
<td>1.076</td>
<td>1.117</td>
<td>1.087</td>
</tr>
<tr>
<td>( GR_{TMNT} )</td>
<td>1.202</td>
<td>1.175</td>
<td>1.187</td>
</tr>
<tr>
<td>( GR_{CTRL} - GR_{TMNT} )</td>
<td>-0.1258</td>
<td>-0.0585</td>
<td>-0.0998</td>
</tr>
<tr>
<td>Deg freedom</td>
<td>337</td>
<td>321</td>
<td>337</td>
</tr>
<tr>
<td>p-value</td>
<td>0.000</td>
<td>0.195</td>
<td>0.001</td>
</tr>
</tbody>
</table>

6.2 Grade Ratio
Grade ratio (GR) scores (described in 5.2.2) were compared between cohorts using two-sample T-tests. Comparisons were made for grade ratios computed using the short-term, long-term, and course total raw academic scores. The results are summarized in Table 3.

6.3 Confidence Scores
Composite confidence (CC) scores (described in 5.2.3) were compared between cohorts using two-sample T-tests. Comparisons were made for the pre-instruction and post-instruction scores. The results are summarized in Table 4.

<table>
<thead>
<tr>
<th></th>
<th>Pre-block</th>
<th>Post-block</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC_{CTRL}</td>
<td>3.665</td>
<td>3.932</td>
</tr>
<tr>
<td>CC_{TMNT}</td>
<td>3.836</td>
<td>4.094</td>
</tr>
<tr>
<td>CC_{CTRL} - CC_{TMNT}</td>
<td>-0.171</td>
<td>-0.162</td>
</tr>
<tr>
<td>Deg freedom</td>
<td>234</td>
<td>255</td>
</tr>
<tr>
<td>p-value</td>
<td>0.099</td>
<td>0.098</td>
</tr>
</tbody>
</table>

7 DISCUSSION
No significant difference was found in raw academic scores. Average scores favored the control group, as expected due to that group’s higher average CQPA. However, effect sizes were small and statistically insignificant (all \( p > 0.1 \)). This confirms we achieved the goal of providing similar instruction and measurement metrics to the two cohorts; neither cohort was disadvantaged. In other words, neither cohort received deliberate, assessment-oriented instruction; both cohorts had similar academic performance.

However, the lower CQPA of the treatment group means that this equivalent performance is, in fact, significant. This is evident in viewing the grade ratio analyses. Both cohorts achieved average grade ratios greater than 1.0, meaning students in either group, on average, performed better than their CQPAs forecasted. However, the average grade ratio was in favor of the treatment group for all assessments. The effect, however, was strongest at the short-term; in the long-term, the two groups were no longer statistically distinct. Both scores show the treatment group performing, on average, better than projected, based on CQPA; and, the treatment group exceeded expectations by a significantly greater margin than the control group. This is strong evidence that the treatment curriculum may be more effective in communicating database concepts and practice, although the effect seems to diminish with time. This diminishment may be explained, in part, due to: student familiarity with the topic by the time the final exam is taken; the final exam being open-note and open-book; or the relatively trivial practical application on the final due to the time constraints students face during the final.

The small effect size in grade ratio, 0.0998, might seem underwhelming. However, we must consider the observed performance difference this indicates. Considering the course total results (reflecting all measured database assessments), on average, students in the control group scored 108.7% of their anticipated score, as projected by their incoming CQPA. We contrast this to students in the treatment group, who scored, on average, 118.7% of their anticipated score. As a practical example, consider an average student with a 3.0 incoming CQPA; his anticipated overall grade and thus for each individual module is a 3.0 equivalent, a B. In the control group, this student
was likely to achieve a 3.261 equivalent grade, just under a B+. In the treatment group, however, this student was likely to achieve a 3.561 grade, almost an A-. In these terms, the significance of the observed effect is clear.

Both groups showed an increase in average reported confidence. The treatment group had, on average, higher confidence scores, both before and after. The results are marginally significant ($p=0.1$), and the effect size is roughly equivalent between groups. This fails to support our hypothesis that the treatment curriculum would provide greater confidence. We expected to see equivalent confidence prior to the block, and significantly greater confidence in the treatment group afterward. The obtained results may indicate a systemic flaw in the measurement methodology for confidence. The results may also fail to account for presentation variation between instructors; it is likely that the treatment instructors were more enthused about the block than the control instructors, and may have imparted undue initial confidence to their students.

8 CONCLUSION AND FUTURE WORK

In this paper, we presented an analysis of a large sample of performance measurements across a control and treatment group with the major graded assessments held constant between them. Our experimental results indicate that those students who were exposed to problem-centered learning versus skills-based instruction significantly outperformed their counterparts relative to their incoming CQPA.

Because our study examined primarily 3rd-year students in the 2nd semester of the academic year, we believe CQPA to be largely indicative of a student’s capabilities. Indeed, our statistical analysis shows this correlation between an incoming CQPA and a final grade in the course. Because the correlation is not perfect it allows for other factors, such as the method of instruction, to contribute to the outcome. Thus, the problem-centered learning method used in our course appears to have improved students’ performance above what was anticipated.

While the volume of work is robust regarding problem-centered learning’s benefits over traditional skills-based instruction for information technology concepts and applications, much of it is domain-specific to information technology, computer science, or traditional business courses. Additionally, a number of the studies relied solely on student feedback or instructor-observations to evaluate the success of the particular pedagogical approach. We demonstrated with statistical significance that the problem-centered learning methodology does improve a student’s ability to perform. This reinforces the notion that problem-centered learning is preferable for information technology topics in courses supporting a student’s cyber education.

As institutions implement core education requirements in the cyber or computing domain as part of their multi-level cyber education programs, curriculum caretakers must be mindful of the impact such a compressed timeline can have on the quality of learning. As demonstrated in this paper, one method for improving the efficient use of the limited time allocated to each topic is to utilize a problem-centered pedagogical approach.

Because of the constraints placed on section distribution amongst teachers in either the control or treatment groups, it is possible that stronger instructors influenced the results. To further demonstrate the benefit of problem-centered learning in cyber topics, future research will attempt to control more strongly for bias amongst instructors. We seek to implement this problem-centered learning methodology across all modules within our course based on these results while continuing to evaluate student performance.

REFERENCES


