2016

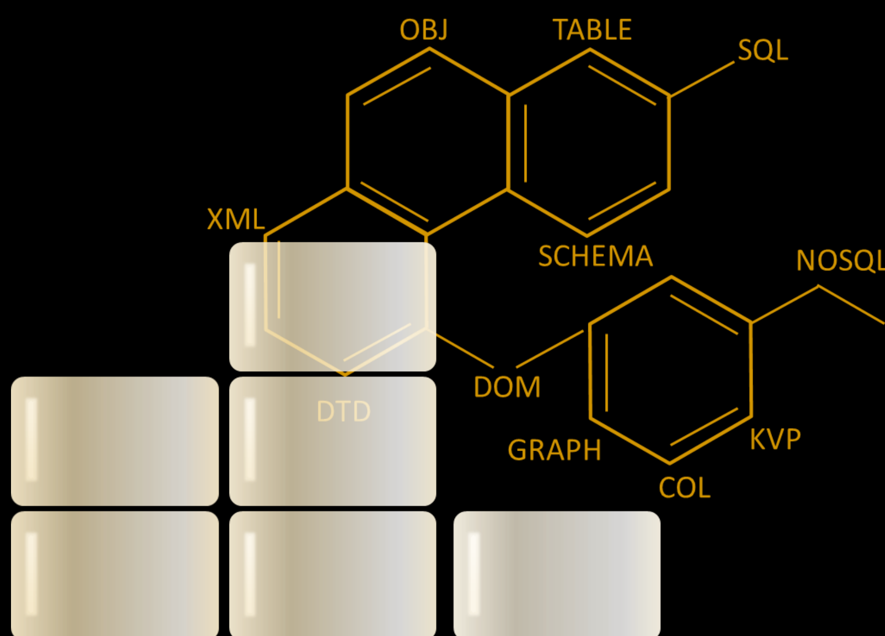
CSCE 411H: Data Modeling for Systems Development—A Peer Review of Teaching Project Inquiry Portfolio

Hongfeng Yu
University of Nebraska-Lincoln, yu@cse.unl.edu

2015-2016

Advanced Peer Review of Teaching Project Inquiry Portfolio

CSCE411H Data Modeling for Systems Development



Hongfeng Yu
Computer Science and Engineering
University of Nebraska-Lincoln

6/6/2016

Abstract

A new course CSCE411H has been developed in 2015-2016. The course tackles the learning of traditional and emerging data modeling techniques in big data related areas from the system and application perspectives. The students have mixed background in Business, Engineering, and Art and Science with different levels. These have introduced a unique set of challenges in the development of this new course. In this inquiry portfolio, I investigated if the adjustment of assignments can benefit the team work of the students with a variety of background. Through the data collection and analysis, the investigation showed that the new assignment design can facilitate the students to reach the learning goals. It also suggested that more effort would be desired to design assignments to help business students in team work with increasing complexities. Although this inquiry portfolio targets a specific question, the general methodology can help us systematically investigate and address other issues in teaching and learning activities.

Keywords: student background, assignment design, team work, survey, teaching data collection and analysis.

Table of Contents

1. Introduction

A new course CSCE411H "Data Modeling for Systems Development" has been developed and taught. Through the development of this course, I identified a few interesting questions for investigation and developed this inquiry portfolio. This portfolio focused on one particular identified question, introduced the methodology used in the investigation and analysis, and discussed the discovery and the possible solutions.

1.1 Course History and Development

With advanced computing techniques, date-driven research and study become ubiquitous across various scientific and business areas. However, the related courses in computer science are comparatively rare. In particular, the core topics on data modeling have not been adequately covered in computer science courses. To this end, in Spring 2014, CSCE378H "Introduction to Data Modeling for Systems Development" has been proposed and designed to prepare students with the knowledge and techniques on data modeling to handle big data related problems. I taught CSCE378H twice in Spring 2014 and Spring 2015. In Fall 2015, given the depth and breadth of the course, CSCE378H has been upgraded to a new course CSCE411H "Data Modeling for Systems Development".

Designing appropriate data models is the key to the success of systems development. The previous CSCE378H covered different levels of data modeling techniques: conceptual data modeling, logical data modeling, and physical data modeling, which are used at the different stages of systems development. At the very beginning stage, technology-independent specific of data needs to be extracted at a conceptual level. It focuses on the communication between data modelers and users. Then, conceptual models need to be translated into logical models that can be implemented using a database/data management system. Finally, a physical model needs to be built according to a specification of physical storage/placement and access mechanism. Therefore, the course covered the whole course of data modeling and relied on several fundamental computer science courses, including programming languages, algorithms, data structures, software engineering, and so on. In addition, the course was application-driven and design-driven, and was closely related to real-world applications.

Based on the coverage of CSCE378H, the new CSCE411H increases both the depth and the breadth of data modeling techniques. Specifically, apart from the traditional data modeling techniques for structured data (e.g., the entity-relationship [E-R] modeling and the corresponding database management systems), the new course introduces the techniques for semi-structured and unstructured data (e.g., graph data and document data) that are widely used in nowadays applications. In addition, the new course covers both the system and application knowledge and techniques of data modeling. From the system perspective, the physical techniques have been introduced for entity-relationship modeling and graph modeling. From the application perspective, students are exposed to the practices to build end-to-end data modeling and analytics systems. Hence, built upon CSCE378H, CSCE411H aims to provide students with more advanced and holistic knowledge and technology on data modeling for real-world big data applications.

1.2 Philosophy

CSCE411H and CSCE378H tackle the topics in the emerging big data related areas, and thereby share the same challenge in that no existing textbooks and courses have appropriately covered the depth and the breadth of the data modeling algorithms and concepts for undergraduate students. In practice, new techniques have been routinely developed and proposed with the increasing complex and scale of applications. This requires that a data modeling course covers not only a particular set of techniques, but also the design knowledge and concepts to deploy these techniques in software systems development. Thus, I use a combination of relevant research papers, documentation from relevant systems, readings from the Web, as well as a set of recommended books. The relevant materials are organized and presented via the slides during the lectures. The slides and the corresponding reading list are provided to the students after each lecture.

In order to close the loops of an entire data process pipeline, I also introduce the topics related to data analysis and visualization techniques. These topics incorporate data analysis and computational methods into systems development, and inspire students to study the interplay between data models and these analytics techniques. In addition, I carefully re-design assignments to facilitate the student learning in CSCE411H. The assignments are completely renovated and organized to cover the scope of CSCE411H. After the complete of each assignment/project, a review session is held to revisit the corresponding content covered in the lecture, and their applications reflected in the assignment. The students are encouraged to discuss during the lectures, and I also prepare questions for students' thinking. Students are encouraged to stop by my office. Multiple students have visited me for discussing topics and questions arising from lectures, assignments, and readings.

I have employed active learning techniques in the classroom of CSCE411H. In particular, I have developed a few games that partition the students into different groups and use the inter-group activities to mimic the process of a software system. For example, to illustrate the concept and the advantage of parallel processing, I ask the students to conduct a set of computation tasks in a sequential fashion and in a divide-and-conquer fashion. By comparing the performance, students are exposed to the difference of these two processing manners. The similar activities have been designed to explore several topics, including graph store, data indexing, data query, and so on. The positive feedbacks on these activities have been received from the students.

2. Identifying an Issue to Investigate

CSCE411H is offered to the undergraduate students of Jeffery S. Raikes School of Computer Science and Management (a.k.a, Raikes School) at UNL with a comparatively high standard of enrollment. The students have the mixed background of Art and Science, Engineering, and Business. One of the main challenges is how to effectively achieve the learning objectives in a way that is not only with appropriate technical depth and breadth, but also is suitable for the students with a variety of background.

2.1 Description of Students

In the spring semester 2015, the total enrollment of CSCE378H is 33 students, in which 1% are sophomore, 82% are junior, and 17% are senior. In terms of background, 42% are from Art and Science, 27% are from Engineering, and 33% are from Business. In the spring semester 2016, the total enrollment of CSCE411H is 25 students, in which 4% are sophomore, 84% are junior, and 12% are senior. In terms of background, 52% are from Art and Science, 40% are from Engineering, and 8% are from Business.

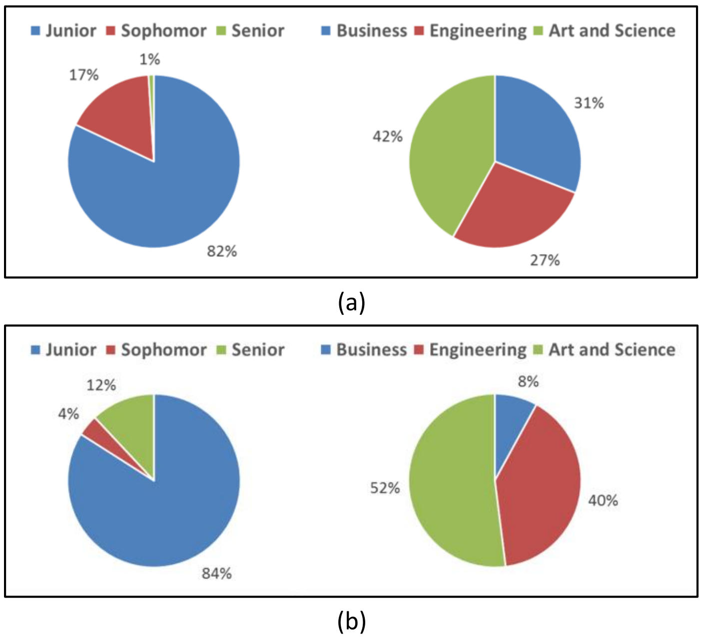
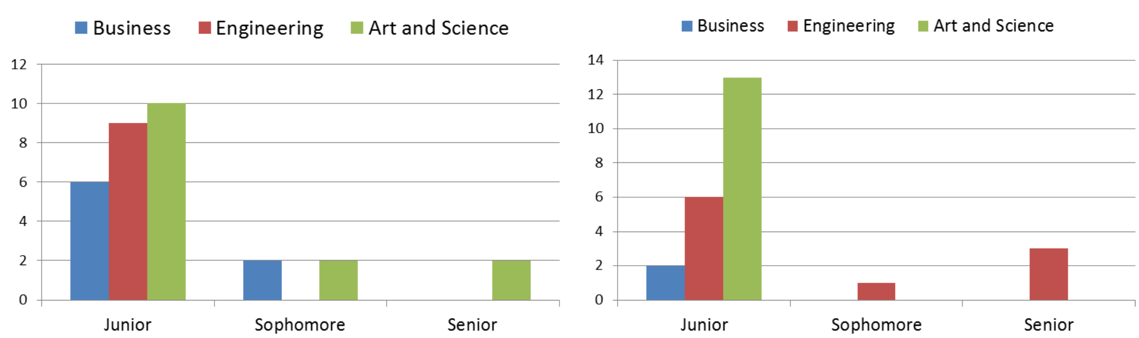


Figure 1: Levels and background of students in Spring 2015 (a) and Spring 2016 (b).



(a) (b)

Figure 2: The numbers of students with different background over different levels in Spring 2015 (a) and Spring 2016 (b).

Figure 1 shows the background and levels of the students, and Figure 2 shows the detailed numbers of students with different background over different levels in Spring 2015 and Spring 2016. We can clearly see that the diversity of the students has been changed considerably over the two years. In general, the majority is the junior students, and the number of students with Art and Science is marginally larger than the other two majors.

2.2 Hypotheses

In the new CSCE411H course, I have covered more advanced topics and changed the assignment design accordingly. In particular, the students are grouped to conduct the team work on each assignment to develop the capability of collaborative systems development. The **central question** I want to investigate is:

- Can we design new assignments and form groups to benefit the team work of my students with a variety of background?

Through the investigation, I aim to study an appropriate assignment design for my students with different background and levels. Meanwhile, such a design can adequately cover the depth and breadth of the new course, meet the learning expectation, and help students understand and apply the subject matter.

3. Methodology

To answer this question, I have designed and conducted surveys in each semester to collect the students' feedback from the assignment design. I also have analyzed if the students reached the designed learning objectives in the assignments. Based on the findings from data collection and analysis, I have refined the assignment design.

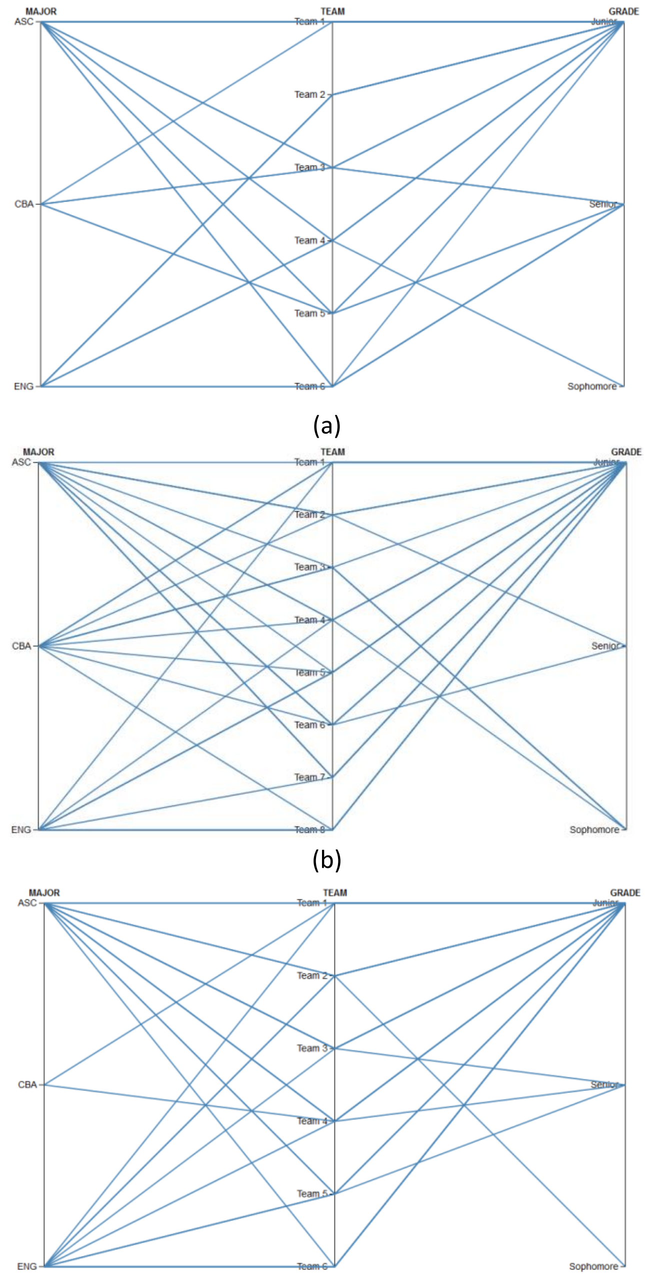
3.1 Team Constitution

In Spring 2014, there were 24 students who were organized into 6 teams, and each team consisted of 4 students. Figure 3 (a) shows the constitution of student teams that is visualized using parallel coordinates. We can see that the student levels were roughly evenly distributed among the teams. However, the major distribution was relatively uneven. For example, the Business students and the Engineering students were assigned among 3 teams, while the Art and Science students were assigned among 5 teams. In addition, we can clearly see that 5 teams have the students from two different majors, and 1 team (Team 2) only contains the students from one major.

In Spring 2015, a pretest was conducted to evaluate the students background before the course. Based on the test results, the 33 students were organized into 8 teams, and each team consisted of roughly 4 or 5 students. Figure 3 (b) shows the constitution of student teams. We can see that the student grades were still evenly distributed among the teams, given the student number in each grade (see Figures 1 and 2). Compared to Figure 3 (a), we can clearly see that the major distribution was more even. The Business students were assigned among 7 teams, the Engineering students were assigned among 5 teams, and the Art and Science students were assigned among 7 teams. Given the student number in each major (see Figures 1 and 2), this constitution shows a more balanced background of students in each team. Specifically, 3 teams

(Teams 1, 4 and 5) consisted of the students from all 3 majors, and the rest teams consisted of the students from 2 majors. No team has the students only from 1 major.

In Spring 2016, the teams were formed using the same method as 2015. The 25 students were organized into 6 teams. As shown in Figure 3 (c), the student background and levels were also evenly distributed among the teams.



(a)

(b)

(c)

Figure 3: Constitution of student teams with different background over different levels in Spring 2014 (a), Spring 2015 (b), and Spring 2016 (c) using parallel coordinates.

3.2 Changing Assignment Design

In Spring 2015, I created two primary assignments. Each assignment focused on different data modeling techniques (specifically, E-R data modeling and graph data modeling). In addition, for each data modeling technique, the number of key algorithms and the size of data were marginal for implementation. The first assignment focused more on the system aspect where the students were asked to implement physical E-R data modeling techniques with B+ tree creation and search operations using small data sets. The second assignment focused more on the application aspect where the students were asked to use graph data modeling techniques for system development, and conduct a comparison study between E-R and graph-based techniques. Therefore, the over assignment design covered the different data modeling techniques from the system and application aspects, as shown in Table 1.

Primary Assignment Design in Spring 2015
Assignment 1: System Implement physical E-R data modeling techniques with B+ tree creation and search operations using small data sets.
Assignment 2: Application Use graph data modeling techniques for system development, and conduct a comparison study between E-R and graph-based techniques.

Table 1: The primary assignment design in Spring 2015.

Primary Assignment Design in Spring 2016
Assignment 1: Application Design and implement an end-to-end E-R data analytics workflow by applying E-R and DOM data modeling techniques.
Assignment 2: System Implement physical E-R data modeling techniques with B+ tree creation and search operations using large data sets.
Assignment 3: System Implement physical unstructured and graph data modeling techniques using small data sets.
Assignment 4: Application Design and implement an end-to-end graph data analytics workflow, and conduct a comparison study between E-R and graph-based techniques.

Table 2: The primary assignment design in Spring 2016.

In Spring 2016, given the new depth and breadth requirements for CSCE411H, I created four primary assignments. Each assignment was also designed for different data modeling techniques, and focused on either the system aspect or the application aspect. In particular, I adjusted the sequence of system and application assignments. I started with an application assignment to inspire the students by designing and implementing an end-to-end E-R data analytics workflow using E-R and web-based data modeling techniques. This assignment was close to the real-world applications that could be more intuitive for the students and help them link the data modeling techniques to the software tools used in our daily life. Then in the second assignment, I led the students to the system details of software tools from the data modeling perspective. This assignment was similar to the first assignment in 2015, but I re-designed it with more complex data sets and more detailed comparison studies. Based on the traditional structured data modeling problem in the first and second assignments, the third assignment was designed to expose the students to the more challenging semi-structured and unstructured data problems. The students were asked to implement physical graph and unstructured data modeling techniques using different data sets. The fourth assignment went back to the application aspect, and asked the students to design and implement an end-to-end graph data analytics workflow, and conduct a comparison study between E-R and graph-based techniques. Therefore, the overall assignment design still covered different data modeling techniques from the system and application aspects, but required more technical depth and involved more boarder applications, as shown in Table 2.

4. Data Collection and Analysis

4.1 Student Survey Results

In order to measure the effectiveness of design methodology, I conducted the surveys of the following six questions[1] related to team work in Spring 2015 and Spring 2016.

Q1: I performed my share of the team's work

1(Strongly Disagree) 2(Disagree) 3(Indifferent) 4(Agree) 5(Strongly Agree)

Q2: I provided relevant and timely information and research to the topic under study

1(Strongly Disagree) 2(Disagree) 3(Indifferent) 4(Agree) 5(Strongly Agree)

Q3: I was cooperative and worked with the group to reach common goals

1(Strongly Disagree) 2(Disagree) 3(Indifferent) 4(Agree) 5(Strongly Agree)

Q4: My partners performed their share of the team's work

1(Strongly Disagree) 2(Disagree) 3(Indifferent) 4(Agree) 5(Strongly Agree)

Q5: My partners provided relevant and timely information and research to the topic under study

1(Strongly Disagree) 2(Disagree) 3(Indifferent) 4(Agree) 5(Strongly Agree)

Q6: My partners were cooperative and worked with the group to reach common goals

1(Strongly Disagree) 2(Disagree) 3(Indifferent) 4(Agree) 5(Strongly Agree)

Figure 4 shows the average score of each survey question in 2015 and 2016. Both show the effective team work in the assignments, while the average scores in 2015 are slightly higher.

[1] Tim Wentz. Course portfolio for ARCH333/CNST305: Building Environment Technical Systems. Advanced Peer Review: Team Formation in a Multi-disciplinary Course Featuring a Service-Learning Project. May 18, 2005

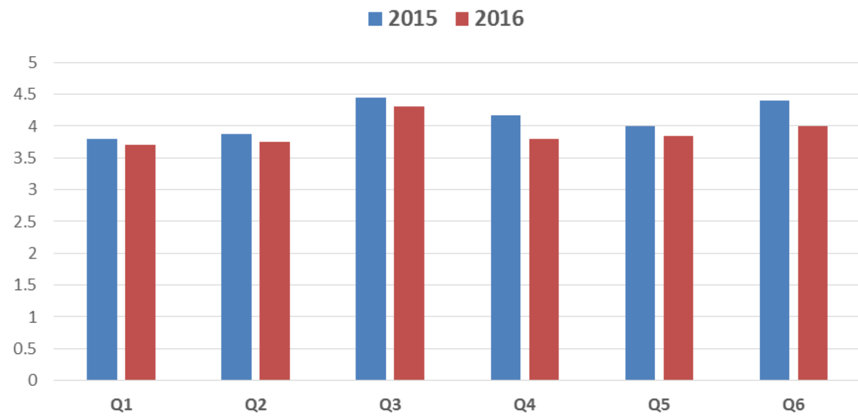


Figure 4: Survey results of the six questions in 2015 and 2016.

In 2015, the survey did not distinguish the background of the students. Thus, the finer results were not captured. In 2016, I asked the students to provide their background information, and the survey results were further broken down accordingly, as shown in Figure 5.

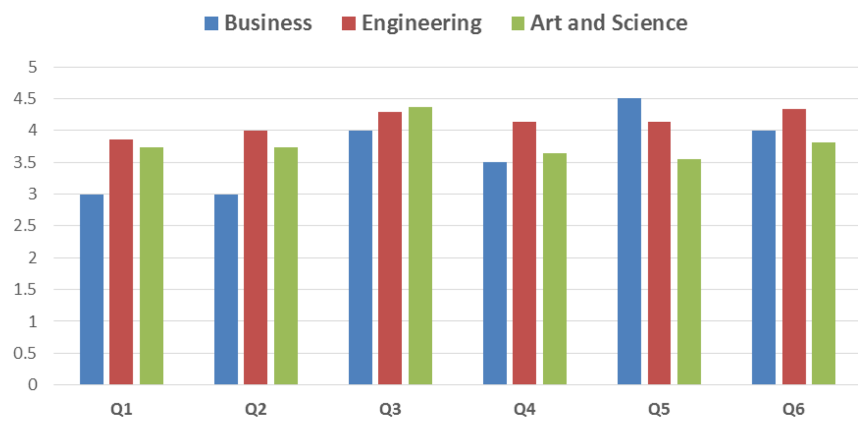


Figure 5: Detailed results with respect to the student background in 2016.

4.2 Analysis

Although the assignment design has been changed, the overall student responses are changed marginally over the two years as shown in Figure 4. However, there are some subtle variations conveyed in Figures 4 and 5.

Figure 4 shows that in the both years the students agreed that they and their partners performed the share of the team's work and provided relevant and timely information and research to the topics under study. We can see the scores of Questions 1/2 and Questions 4/5 are coherent. Moreover, each student provided a relatively high score to Questions 3 and 6, expressing that she/he and the partners were cooperative to reach the learning goals. The scores of Questions 3 and 6 are consistent.

However, we note that each score in 2016 is slightly lower than that in 2015. Thanks to the results in Figure 5, more details can be perceived to understand the difference.

In Figure 5, from Questions 1 and 2, we can see that both the students of Engineering and Art and Science were more confident on their own performance. They felt that they were able to perform their share of the work and provide useful assistances in team work. However, we can clearly see that the business students gave themselves significantly lower scores. They were less confident on their share of the team's work.

Correspondingly, in Question 5, the business students almost strongly agreed that their partners provided relevant and timely information and research during the work. The score was 4.5, significantly higher than the ones of Question 5 in Figure 4.

From Question 4, we can see that the engineering students were more comfortable with their partners and the score was higher than 4. However, the rest students of Art and Science and Business provided relatively lower scores to their partners. In general, Questions 1/2 and Questions 3/4 have the consistent scores.

From Questions 3 and 6, we can see that the business and engineering students had a balanced score for themselves and their partners. The students from Art and Science showed a slightly different opinion, and rated themselves more cooperative than their partners.

4.3 Reflection

From the overall responses, the survey results indicated that the change of assignment design has slightly affected on the effectiveness of team work. In general, the students agreed that they and their partners can work collaboratively to solve the problems and apply the knowledge to reach the learning goals.

However, by examining the survey results according to the student background, more information can be revealed. We can see the business students gave comparably lower self-rating scores, while the scores of engineering and science students were more balanced.

Revisiting the hypotheses, we can derive the following answers.

Central question: Can we design new assignments and form groups to benefit the team work of my students with a variety of background?

Answer: Yes, we can design assignments that can facilitate the team work of students with mixed background across Business, Engineering, and Art and Science. Although the number and complexity of the assignments have been increased for the new course, the students can leverage the combination of the assignments to reinforce the learning and application of the knowledge in a collaborative fashion. In general, the students agreed that they and their partners can provide the needed information and conduct the share of the work in a collaborative fashion. As a team work, they can reach the common learning goals in the assignments.

In addition, we note that more effort is needed to design assignments to help non-engineering students in team work. Thanks to the new design of the survey, we can distinguish the responses of the students with different background. We can clearly see that the business students were not best to perform their share of the team work and reply more on the assistance from their partners. They strongly agreed that their partners were helpful. As we already assigned the businesses students to the different teams with balanced background, this suggests that a possible adjustment of the assignments would be needed. One possible adjustment is that, apart from the system and application aspects, more non-technical or business factors could be incorporated into the assignments. Examples include software product design and user studies, which are important components in a life cycle of holistic software system design, but have not been explicitly reflected in the assignments yet. I will develop corresponding solutions and collect data, including peer reviews and student performance and feedbacks, to evaluate and refine the solutions. Given the increasing requirement from non-engineering students to seek computer science courses, I expect that the diverse background of students would become a common challenge in the future teaching.

5. Future Inquiry and Development

Through the development of this inquiry portfolio, I gained more experiences in addressing challenges in developing a new course. I have explored assignment design to tackle the new depth and breadth requirement of the course. The survey results demonstrated the assignment adjustment can still facilitate students' team work and learning activities.

In addition, by conducting a finer-grained survey with respect to the student background, I can obtain a deeper understanding of student learning. More detailed assignment adjustment would be needed to facilitate the business students on the team work. In the future, I plan to include more business factors into the assignment and enhance the students' learning on software systems development. Correspondingly, I will improve the survey to capture the students' responses reflecting the effectiveness of the changes.

Given the fast evolution of computing techniques, it could be a nature need for an instructor to develop courses and teaching techniques to meet the new emerging requirements. Although this inquire portfolio was developed to address a specific question in teaching a new course on data modeling, I enhanced experiences on the general methodology and philosophy for solving problems. These experiences and lessons will be helpful for me and other teachers to systematically investigate and address other emerging issues in teaching and learning.