TEACHING DATA SCIENCE TO POST GRADUATE STUDENTS: A PRELIMINARY STUDY USING A "F-L-I-P" CLASS ROOM APPROACH

Sunet Eybers and Mariè Hattingh
University of Pretoria, Private Bag X20, Hatfield, South Africa

ABSTRACT

Data is everywhere. As a result the need for data scientists with the correct skill set to analyze and interpret the data has escalated. Not surprisingly, data scientists are currently one of the most wanted professions. Tertiary institutions are faced with the challenge of producing students with the correct blend of theoretical knowledge and practical skills. In an attempt to provide current post-graduate students with these skills, a flipped class room approach was adopted to teach students data warehousing as part of the data science curriculum. This paper used the four pillars of F-L-I-P (FLN 2014) to analyze data obtained from lecturer implementation experience and student course evaluation forms. It was shown that great strides have been made in this course in the adoption of the flipped class room although there is room for improvement in order to achieve flipped learning. The outcome of the study directly contributes to the improvement of subsequent implementations of the module.

1. INTRODUCTION

Data science is a relatively novel discipline (Song & Zhu 2015; Provost & Fawcett 2013) born as a result of the need to capitalize on vast volumes of data, also referred to as "big data". It is a combination of older disciplines such as mathematics and statistics as well as the newer discipline of computer science (Provost & Fawcett 2013). The multidisciplinary nature of data science has created a challenge to educational institutions to provide students with the necessary blend of knowledge and skills to fill the anticipated need for competent data scientists (Provost & Fawcett 2013). The unique blend of knowledge and skills are often challenging to teach in a traditional class room environment (Turek et al. 2015).

In an attempt to address the challenges associated with the traditional class room environment a blended learning approach is suggested. Blended learning is a method whereby different types of education methodologies and technologies are adopted to provide more "effective education experiences" (Köse 2010). As a result, a combination of older traditional face-to-face models and online technologies are utilized in the teaching model.

The flipped class room approach is one of many blended learning approaches. Bishop and Verleger (2013) describes this method as *"a new pedagogical method, which employs asynchronous video lectures and practice problems as homework, and active, group-based problem solving activities in the class room"*. Therefore, the traditional activities conducted in the class room and at home is switched (or "flipped" / "inverted"), i.e. lectures are watched at home whilst homework questions are discussed in class (Herreid & Schiller 2013; Bishop & Verleger 2013). A flipped class room approach was adopted in the teaching of a post graduate data warehouse semester course under the umbrella discipline of data science. A flipped class room approach was adopted for three main reasons:- 1) to improve the learning experience of students; 2) face-to-face contact time was limited (one contact session every fortnight) and the total number of concepts could not be covered in detail during these sessions; and 3) the course was presented on post-graduate level where the requirement was that students should master a higher level of thinking skills than the traditional recall type testing. A flipped class room approach seemed more applicable to provide a more in-depth presentation of concepts.

The main objective of the study was to compare student evaluations and lecturer implementation experience of traditional teaching methods (2013 and 2014 cohort) with a flipped class room approach (2015 and 2016 cohorts). As this was the first time that a full flipped class room approach was adopted in the course (as opposed to a partial flipped class approach in 2015 and traditional approach in 2013 and 2014), the main aim was to evaluate the extent to which flipped learning was achieved by using the four pillars of the F-L-I-P (FLN 2014). The anticipated contribution is to identify areas of success and improvement of the 2016 cohort. This can contribute to the successful implementation of the teaching of data science related subjects in general.

The outline of the paper is as follow: the four pillars of F-L-I-P is described followed by a discussion on the approach in the context of data science. The methodology and data is presented followed by a discussion and conclusion.

2. THE FOUR PILLARS OF F-L-I-P

Further to the brief definition of the flipped class room provided above, the four pillars of F-L-I-P is one of many pedagogical models developed to support educators to transition from a traditional class room to a flipped class room. Examples of models included ATRACT (Coley 2012) which describes issues to be aware of when considering to implement a flipped class room and CPBL (Warter-Peres and Doug 2012) that suggest that learning should take place outside the class room which will then allow learners to complete problem based learning activities in the class room assisted by their peers and educators.

Flipped Learning Network (FLN 2014) defined the four pillars of flipped learning in order to make the very important distinction between FLIP class room and FLIP learning. It is argued that although a number of educators might employ the FLIP class room approach by giving students preparatory work, it is not clear whether any learning takes place. FLIP learning is defined as *"a pedagogical approach in which direct instruction moves from the group learning space to the individual learning space, and the resulting group space is transformed into a dynamic, interactive learning environment where the educator guides students as they apply concepts and engage creatively in the subject matter"* (FLN 2014:2). In order to ensure that learning takes place in a flipped class room they propose four pillars of F-L-I-P where each pillar has a number of specific objectives to be met. Each of these pillars will be briefly introduced in turn. These pillars will be used in section 5 as the analytical lens through which the data will be discussed.

2.1 Flexible Environment

The "flexible environment' pillar refers to both the physical as well as the virtual student engagement space. It caters for both individual and group interactions. The flexibility of the pillar refers to the anytime and anywhere access and engagement with the learning environment. A flexible learning environment can be enhanced by technology, for example web 2.0 tools. These tools can *"encourage active learning, promote collaboration, increase student-faculty interaction and enrich educational experience"* (Mandernach & Taylor 2010). Students can therefore select "when and where" they learn. From a lecturer perspective the expectation of student timelines for learning and assessments is flexible as long as key deadlines are met.

2.2 Learning Culture

Learning culture refers to moving away from the traditional teacher-centered approach where the teacher is the primary source of information to a learner centered approach where in-class contact time is utilized to discuss topics in greater detail therefore creating a richer learning opportunity. As a consequence students are *"actively involved in knowledge construction as they participate in and evaluate their learning in a manner that is personally meaningful"* (FLN 2014).

2.3 Intentional Content

Intentional content refers to the learning material to be used to cultivate the student's conceptual understanding as well as practical application of concepts. Facilitators or lecturers should therefore carefully consider both the content as well as presentation method of material prior to contact sessions to maximize the

effectiveness of these sessions. The objective should be on achieving the best possible learning experience for the student by a combination of student self-discovery and theoretical material.

2.4 Professional Educator

The educator in a flipped class room can be compared to a "silent partner" and facilitator. As a result the role of the lecturer can be much more demanding when compared to a traditional class room setup. Furthermore the lecturer requires a more specialized skill set than in a traditional class room. Continuous feedback by the educator is of utmost importance in the learning process. Contact sessions are often characterized by "chaos" but with constructive feedback.

3. FLIPPED CLASS ROOM FOR THE DATA SCIENTIST

Earlier research in the field of science, technology, engineering and mathematics (STEM), of which data science is a subset of, has focused on the implementation of the flipped class room approach for teaching STEM.

In a study conducted by Fulton (2012) based on teaching Calculus 1, it was found that the flipped class room approach as a positive allowed: (1) Students the opportunity to study at their own pace; (2) Students who experienced difficulties to be identified and assisted during the learning process; (3) The curriculum to be adapted to address difficult concepts according to the student's needs; (4) Contact-time to be used in a constructive way; and (5) Students to be more interested in the subject. However, as a negative, students new to the concept of the flipped class room approach perceived the "homework" as additional work. As a result, not all the students conducted the additional preparatory homework assignments. Homework had to be developed and customized to the level of the student. As a result, this approach was time intensive for the lecturer.

The content of data science curriculums are a topic of much discussion amongst academics (Song & Zhu 2015; Tang & Sae-Lim 2016; Yi 2016; Dubey & Gunasekaran 2015). Whilst the focus of data science curriculum related studies are mainly on what should be included, a few studies have now started to focus on how data science related subjects should be taught. Turek, Suen and Clark (2015) argued that in order to prepare data scientists for "the real world" data science education should adopt a multi-disciplinary approach based on experiential learning (flipped class room) methodology. They have called for three considerations in designing data science courses: Firstly, "real-world projects" should be used in order to present students with accessible data and in a real world context, secondly team dynamics on a project level was surprisingly an important factor in the successful operation of the research team and finally early introduction of technical tools is recommended in order to ensure effective team dynamics. As a result a multi-disciplinary approach to data science education is proposed in order to deal with both the technical and team dynamics.

A study conducted by Dichev et al. (2016) focused on adopting an active participatory learning approach in contrast with the passive learning approach to teaching an introductory data science course. They agreed with Turek, Suen and Clark (2015) that a learner centered approach which they called a "flipped learning model" is needed to cater for the multidisciplinary requirements of the subject. A phased approach is suggested in developing the data science course.

4. DATA AND METHODOLOGY

One of the courses presented as part of the data science education path at a tertiary institution in South Africa is data warehousing. This semester module was presented to post graduate students (4th year level) who complied with the pre-requisite of completing any relevant undergraduate database-related subject.

The main objective of the data warehouse semester module was to introduce students to the concept of data warehousing that supports business intelligence (BI) in an organizational environment. Although the emphasis was on basic data warehousing concepts without disregarding the influence of big data and disruptive technologies currently changing the landscape. Despite these technologies the fundamentals of

data warehousing and business intelligence remain relevant and provided the departure point to acquire the necessary knowledge to work with structured data.

The course covered six topics namely an overview of the concepts of data warehousing and business intelligence (BI); the planning, design, architecture and infrastructure of data warehousing; data design and preparation focusing on dimensional modeling as well as the loading of data (extract, transform and loading) and data quality; information delivery and online analytical processing (OLAP); and finally the implementation, deployment and maintenance of the environment. For each of the topics, a theoretical section introduced the student to the topic followed by a practical component whereby students had the opportunity to apply the theoretical concepts. The course was presented to a diverse student group. The student groups were diverse both in terms of demographics and academic background. A total of 55 students participated in the course of which 34 were males and 21 females.

An open book approach was adopted. Students were permitted to have access to any of the study materials available. No test recall type questions were asked during assessments. The approach was as follow: one day prior to the class contact sessions, scheduled every fortnight, students had to complete a preparation assignment using the online learning environment. This preparation was based on the material to be covered during the subsequent class contact session. In order for students to complete the assignment successfully they had to work through the respective chapters in the prescribed book as well as watching general videos on the topic available through the learning management system. Students were encouraged to complete these assignments in order to prepare them for the class discussions during the contact sessions. A total of four assignments were available which contributed a substantial amount towards the final semester mark.

The data used in this study was collected from anonymous student evaluations after the course of the cohort 2013 and 2014 which followed a traditional teaching approach as well as a partially flipped class approach in 2015 and a full flipped class room approach in 2016. The evaluation sheets allowed students to score their experience of the module on a Likert scale where 1 is poor and 5 is excellent. There was also ample provision for students to provide written feedback on each of the evaluation areas. These evaluation areas included module content (organization, study material, usefulness of material and usefulness of the course), lecturers (interpersonal relationships, level of knowledge, attitude, preparation and use of media as well as learning opportunities) and assessment (clarity of criteria, nature, content and method as well as fairness), level of difficulty, workload and provision for free text general comments. Table 1 contains a summary of the individual scores per category for the different cohorts. The class contact sessions followed the following structure: key theoretical concepts were covered in class by the lecturer where after the group class discussions followed to practically apply the concepts covered.

The assessment structure for all the cohorts of the module included an semester test, a practical project as well as a final written examination. For the practical project students were provided with a number of data files. Based on a real life business case and business problem students had to use these data files to construct a data warehouse environment containing the data in such a way that the data is available for analysis. The students then had to solve the identified business problem and suggest recommendations by presenting their findings using a presentation layer (graphs / reports / analytical structures).

Table 1. Summary of student evaluation scores

	Average score			
	2013	**2014**	**2015**	**2016**
Area of evaluation				
Module content	**4.29**	**4.00**	**4.04**	**3.73**
Organization (of the course)	4.23	3.85	3.98	3.92
Study material	4.32	4.15	4.14	3.92
Usefulness of material	4.23	3.85	3.90	3.46
Usefulness of course	4.38	4.15	4.15	3.62
Lecturers	**4.43**	**3.99**	**4.28**	**4.03**
Interpersonal relationships	4.58	4.33	4.10	4.08
Level of knowledge	4.51	4.11	4.17	4.23
Attitude	4.6	4.41	4.29	4.23
Preparation, use of media	4.28	3.52	4.31	3.77
Learning opportunities	4.13	3.59	4.52	3.85

	Average score			
	2013	**2014**	**2015**	**2016**
Area of evaluation				
Assessment	**4.02**	**3.87**	**4.14**	**3.82**
Clarity of criteria	3.92	3.81	4.02	3.69
Nature, content and method	3.94	3.85	4.07	3.77
Fairness	4.20	3.95	4.33	4.00
OVERALL AVERAGE	**4.28**	**3.97**	**4.16**	**3.88**

The institutional acceptable average of a 4th year level course evaluation is 3.6. Important to note is that none of the course evaluations specifically references the traditional or flipped class room methodology. Therefore the student responses are indicative of their overall experience of the methodology.

The difference between the average scores obtained for the cohort 2013 and 2014 were vastly different. This was interesting as the course was presented by the same lecturer using the same material and teaching approach (traditional). A possible explanation might be the diversity of the student group for the two years in question. However, this would have to be further investigated. Surprisingly the average score obtained in 2015, when the flipped class room approach was adopted for the first time, was higher than the score obtained in 2016. Further investigation showed that this could be attributed to the fact that a "partial" flipped class room approach was adopted in 2015. For example, in 2015 students were required to read the allocated chapters prior to the class contact sessions. Class discussions were then based on the material covered. In 2016 students were also required to read the allocated chapters but then had to complete online questions and tasks.

For the purpose of this study the student feedback on the module content is of specific interest. In 2016 the flipped class room approach was fully implemented whilst in 2015 it was partially implemented (similar to the phased approach by Dichiev et al. 2016). This might explain the decrease in the overall student evaluation average (from 4.04 to 3.73).

The variance in scores for the lecturer section is very subjective. For every cohort a different student group with diverse demographics participated in the research. However, the same lecturer presented the course in 2013 and 2014. The appointed lecturer changed in 2015 and again in 2016. A possible explanation of the variance might be the different personality and presentation styles of lectures.

For the cohort years of 2013 to 2015 the assessment structures were the same. However, for the cohort 2016 online pre-class assessments were included in the assessments structure. This implied that there was an increase in individual workload prior to class contact sessions. Students indicated in the general section of the course evaluation feedback that the workload was intense therefore scoring the assessment section lower.

5. DISCUSSION

The data obtained from the various cohorts are discussed using the different pillars of the F-L-I-P learning approach (see section 2). This approach fitted well with the categories evaluated on the student evaluation feedback sheets namely module content (intentional content of the F-L-I-P learning approach), lecturers (professional educator of the F-L-I-P learning approach) and assessments (flexible environment and learning culture of the F-L-I-P learning approach).

5.1 Flexible Environment

According to FLN (2014) a flexible environment is achieved by meeting three objectives:
Establish spaces and time frames that permit students to interact and reflect on their learning as needed. In the flipped class room approach (2016 cohort) the lecturers tasked students to conduct activities prior to face-to-face contact sessions by performing online preparatory homework assignments. The learning process was further substantiated by online discussions using a learning management system, online videos and discussions during class contact sessions. Despite the effort by lecturers students felt that the learning opportunities created when a full flipped class room approach was followed was not adequate enough (2016 cohort) when compared to the previous year when a partial flipped class room approach was followed.

Student feedback in cohort 2014 requested more assessment opportunities as well as more supportive media such as videos.

Continually observe and monitor students to make adjustments as appropriate.

The feedback received from the grade system from the preparatory assignments prior to the class contact sessions were used as mechanism to establish the level of knowledge of students. This was used to prepare material for in class contact sessions.

Provide students with different ways to learn content and demonstrate mastery.

Online activities, class-contact sessions and a final practical project provided students with various ways to demonstrate that the required level of knowledge and skills has been achieved. This is similar to the approach adopted by Turek, Suen and Clark (2015).

5.2 Learning Culture

According the FLN (2014) a learning culture is achieved by achieving the following two objectives:

Provide students with the opportunity to engage in meaningful opportunities without the teacher being central.

One of the many advantages of a flipped class room approach is the ability to use contact time more effectively and creatively. However, the students perceived pre-class preparatory assignments as additional work adding to their workload. This is a typical challenge identified by Herreid and Schiller (2013) of students that are new to this concept. Often student resistance is experienced by students new to this approach. This was therefore in line with the responses received of this study. The online management system provided students with the opportunity to engage with peers on specific topics established by the lecturer as illustrated by figure 1. Although the initial topic was established by the lecturer self-propelled student engagement was encouraged.
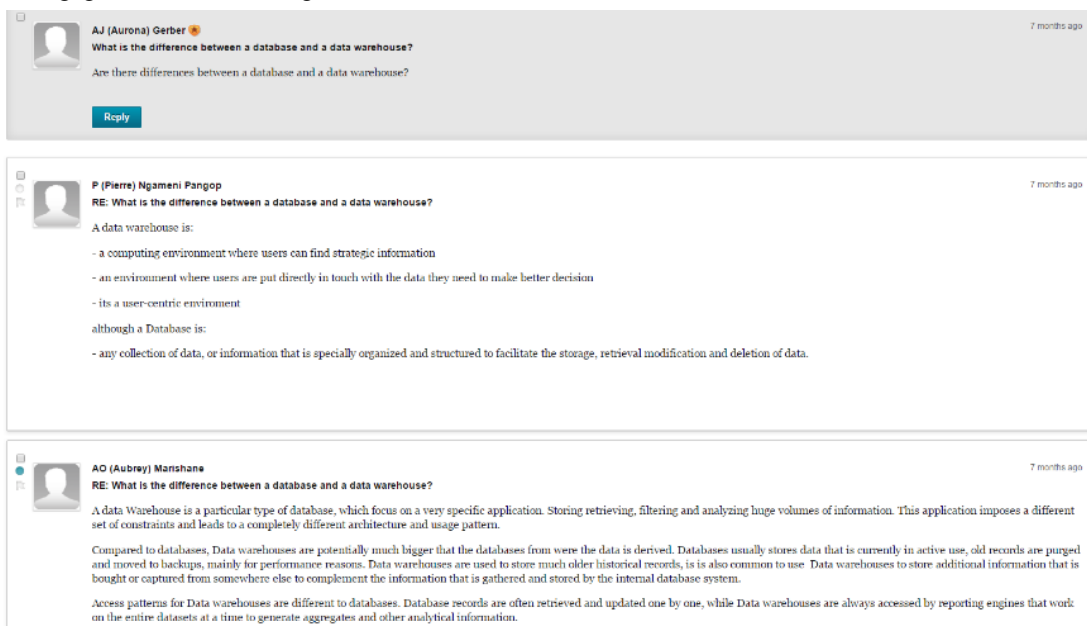

Figure 1. Example of discussion forum

Scaffold these activity and make then accessible to students through differentiation and feedback.

Due to the complexity of the topic (data warehousing) and new technical concepts introduced it was imperative for discussions to complement one another in order for students to conceptualize the new domain. For example the first topic for discussion was the difference between a database and a data warehouse (see figure 1 for example). Consequently the lecturer pitched the class discussion according to her observation of the student's understanding of the concept. The second activity was the need for data warehousing. This activity complimented the first activity by extending the understanding of a data warehouse and at the same time contextualizing the need for this module.

5.3 Intentional Content

Feedback from the student evaluation sheets scored the module content area at 3.73 (2016). This is higher than the minimum average required but the lowest for all the cohorts under consideration. This might indicate that the module content for a flipped classroom needs further investigation to establish what is expected.

According the FLN (2014) a learning culture is achieved by achieving the following three objectives:

Prioritize concepts used in direct instruction for learners to access on their own.

The meaning of the term "data science" is a much debated topic amongst both practitioners and academics (Provost & Fawcett 2013). This can be attributed to the fact that the term is relatively new borrowing concepts from various disciplines (Song & Zhu 2015). The concept also includes new, ill-defined concepts such as big-data and data analytics. Some authors perceive data science as part of the overall concept of "Business Intelligence and Analytics" (BI&A) with a strong focus on the business application of data analytics (Chen, Chiang & Storey 2012). As a result of this ongoing discourse lecturers had great difficulty in identifying and prioritizing of key concepts in preparing students for the "real world".

Create and/or curate relevant content (typically videos) for students.

One of the requirements to implement a flipped class room approach was the customization of preparatory work (Herreid & Schiller 2013). A great deal of work went into the development of online preparatory assignments. This has created an additional workload on the lecturers. But despite the effort, students did not perceive the module content to be extremely useful and organized. However, the module met the expectations of the majority of students. General course videos on the sub-topics were made available on the learning management system. This allowed students to gain an initial understanding of the concepts covered. It also contextualized the concepts using real world examples.

Differentiate to make content accessible and relevant to all students.

Although students did not mention the challenges associated with accessibility of online material, lecturers take cognizance of the fact that this might be a problem in a country such as South Africa, frequently characterized by power outages and high data costs.

5.4 Professional Educator

Feedback from the student evaluation sheets scored the lecturer area at 4.03 (2016) which is second lowest of the cohorts under consideration. Although the student feedback indicated that the lecturer was "knowledgeable" and "professional" further research is needed to understand the student's perception of the lecturer's role in a flipped class room.

According the FLN (2014) a learning culture is achieved by meeting the following three objectives:

Make myself available to all students for individual, small group, and class feedback in real time as needed.

At the start of the semester the learning management system was used to communicate pre-arranged consultation hours to students. Lecturers could also be contacted at any time using online media and additional consultation sessions was available on request. The online discussion forum was monitored and feedback provided where applicable.

Conduct ongoing formative assessments during class time through observation and by recording data to inform future instruction.

Formative assessments were conducted by providing ample opportunity for groups to discuss topics amongst themselves where after one group spokesperson had the opportunity to provide consolidated feedback to the entire class.

Collaborate and reflect with other educators and take responsibility for transforming practice.

This was one of the first modules in the department to follow this approach. The approach followed in 2016 was based on the partial flipped class room approach adopted by the lecturers in 2015. Future collaboration sessions are planned to improve on the implementation of the approach in 2017.

6. CONCLUSION

In an attempt to equip potential data scientists with the correct blend of practical and academic knowledge and skills a flipped class room approach was adopted in a post graduate data science course. Subsequently, the aim of this study was to determine to what extent flipped learning was achieved.

The study compared student evaluations and implementation experience of traditional teaching methods (2013 and 2014 cohort) with a flipped class room approach (2015 and 2016 cohorts). The four pillars of F-L-I-P was used as analytical lens and explains the extent to which flipped learning was achieved.

The findings suggested that great strides have been made in the implementation of a flipped class room approach in particular a flexible learning environment as an online learning management system have been utilized by other post graduate modules. The challenge was that the learning culture of the students still has some adjustment to be made. This is not surprising as it supports previous findings.

The findings of the study further indicated that there is scope for improvement on the provision of customized content in the context of South Africa. The study further suggested that ongoing formative assessments perhaps required more attention as well as collaboration efforts with other lecturers.

The study can be seen as exploratory as the data obtained for the study was based on standard student evaluation feedback forms. Further research incorporating learning analytics and in-depth interviews with learners will be conducted to get a more informed understanding of the effectiveness of the flipped class room approach.

REFERENCES

Bishop, J.L. & Verleger, M.A., 2013, 'The Flipped Class room: A Survey of the Research', 120th ASEE Annual Conference & Exposition, Atlanta, June 23-26, 2013.

Chen, H., Chiang, R.H.L. & Storey, V.C., 2012, 'From Big Data to Big Impact', *MIS Quarterly* 36(4):1165-118. December 2012.

Coley, K.S., 2012, Active learning techniques in the Christian education class room and in ministry context, Christian Educational Journal, 9(2):357-371.

Dichiev, C., Dicheva, D., Cassel, L., Goelman, D. & Posner, M., 2016. Preparing all students for the data driven world, ADMI 2016, March 31-April 3 2016, Winston Salem, USA.

Dubey, R. & Gunasekaran, A., 2015, 'Education and training for successful career in Big Data and Business Analytics', *Industrial and Commercial Training* 47(4):174 – 181.

Flipped Learning Network (FLN), 2014, The Four Pillar of F-L-I-P., online. Available: http://blogs.gre.ac.uk/pcnb/files/2015/07/FLIP_handout_FNL.pdf

Fulton, K., 2012, 'Upside down and inside out: Flip your class room to improve student learning', *Learning & Leading with Technology* 39(8):12-17.

Herreid, C.F. & Schiller, N.A., 2013, 'Case Studies and the Flipped Class room', Journal of College Science Teaching 42(5):62-66.

Köse, U., 2010, 'A blended learning model supported with Web 2.0 technologies', *Procedia – Social and Behavioral Sciences* 2(2):2794-2802.

Mandernach, B. J., & Taylor, S. S. 2010, Web 2.0 applications to foster student engagement. Promoting student engagement, 1, 220-229

Provost, F. & Fawcett, T., 2013, 'Data Science and its relationship to big data and data-driven decision making', *Big Data* 1(1):51-59.

Song, I.Y. & Y. Zhu, 2015, 'Big data and data science: What should we teach?', *Expert Systems*, October 2015.

Tang, R., & Sae-Lim, W., 2016, 'Data science programs in U.S. higher education: An exploratory content analysis of program description, curriculum structure, and course focus', *Education for Information* 32(3):269-290.

Turek, D., Suen, A., & Clark, D., 2015, Pushing data science education into the real world, Bloomberg data for good exchange conference, September 28 2015, New York City, New York, USA.

Warter-Perez, N., & Dong, J., 2012, Flipping the class room: How to embed inquiry and design projects into a digital engineering lecture. In Proceedings of the 2012 ASEE PSW Section Conference.

Yi, M., 2016, 'A Study on the Curriculums of Data Science', *Journal of the Korean BIBLIA Society for library and Information Science* 27(1):263-290.