

Introducing Undergraduate Database Students to K-12 Education Research

Chris Mayfield
James Madison University
Harrisonburg, Virginia, USA
mayfiecs@jmu.edu

Bethann Canada
Virginia Department of Education
Richmond, Virginia, USA
bethann.canada@doe.virginia.gov

Carole Ottenheimer
Center for Innovative Technology
Herndon, Virginia, USA
carole.ottenheimer@cit.org

Brooke Bell
Center for Innovative Technology
Herndon, Virginia, USA
brooke.bell@cit.org

ABSTRACT

Apps4VA is a bold initiative to improve K-12 education by leveraging data on Virginia public school students' progress toward college and career readiness. The goal of the project is for students, professionals, and the general public to create “apps” that explore new ways of using K-12 education data to aid in better data-driven decisions and intervention strategies. In Spring 2013, sixty students in a database systems course participated in a semester-long project based on Apps4VA. This paper describes our teaching experiences, reports our lessons learned working with education data, and documents improvements in student outcomes. We also describe how instructors of database and data analytics courses may partner with local governments to adopt similar projects in their communities.

Categories and Subject Descriptors

K.3.2 [Computer and Information Science Education]:
Computer Science Education

General Terms

Management, Verification

Keywords

Project-based learning, peer assessment, longitudinal data

1. INTRODUCTION

Since the advent of No Child Left Behind (NCLB) in 2001, the role of performance-based standards in K-12 education in the United States has greatly increased. Departments of Education—and ultimately individual teachers—are under increasing pressure to demonstrate student success via

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
SIGCSE'14, March 5–8, 2014, Atlanta, GA, USA.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-2605-6/14/03 ...\$15.00.

<http://dx.doi.org/10.1145/2538862.2538957>.

measurable outcomes, including standardized test scores and graduation rates [1]. A recent study of rural school systems details the challenges and pressures that principals experience in fulfilling NCLB. Recruiting highly qualified, committed teachers as well as “ensuring that all students achieve at levels mandated by NCLB” is overwhelming [10].

The US Congress passed additional legislation in 2009 under The American Recovery and Reinvestment Act (ARRA) that allocated billions of dollars of funding for states to make improvements to their education systems. One of the selection criteria was “Data Systems to Support Instruction,” which resulted in a number of states creating or enhancing systems to track the performance of individual students over the course of their entire K-12 education [6]. The intent was that by using selected sets from the wide array of publicly available data, educators and policy makers will be better able to make informed decisions to improve the efficiency and effectiveness of their schools.

Notwithstanding the controversies fomented by these requirements and systems [8], they are creating new opportunities for leveraging computing to make a positive impact on education at all levels [11]. State departments of education are collecting and releasing an ever increasing amount of data. As computer science educators, we can take advantage of these new sources of data to engage and motivate our students through projects that are relevant to them. After all, today's university students are among the first generation of K-12 students that modern “Data Systems to Support Instruction” were designed to assess.

This paper describes our experiences integrating education research into an undergraduate database systems course. Over the course of a semester, sixty students participated in a coordinated effort to create new applications featuring publicly available data for education researchers. For many of the students, the data was about them, their schools, and their state. We observed a substantial increase in student engagement and performance as a result of this work. In addition, the students gained valuable insight into how large amounts of data may be used to help improve the K-12 education system.

1.1 Apps4VA Initiative

In 2009, the Virginia Department of Education (VDOE) developed a data system that would include linking teachers to their students (while protecting student privacy and

confidentiality) known as the Virginia Longitudinal Data System (VLDS).¹ This system, supported through a federal grant funded through the ARRA program, was created to provide educators and policy makers aggregate data on Virginia school students' progress toward college and career readiness. In order to raise awareness of the existence of VLDS and to highlight its power and potential, VDOE, in collaboration with the Center for Innovative Technology's (CIT) CIT Connect,² launched a software application development program known as Apps4VA. This program was the first apps competition sponsored by a state education department. It challenged the public to review K-12 data and use it to design software solutions ("apps") that would be new and innovative tools to aid in better data-driven decisions and intervention strategies.

A successful apps challenge engages a wide demographic of participants, many of whom might not normally be engaged in the education arena. To that end, in addition to employing traditional methods of outreach, Apps4VA undertook a number of unconventional events and programs to promote community engagement and to reach targeted audiences³ including: a weekend-long startup challenge, four simultaneous hackathons,⁴ a competition for Virginia public high school students, and an open competition for the general public. Participants in all events were encouraged to identify educational needs and to develop software applications to address these needs.

The only development limitation placed on participants was a requirement that these applications would be based on at least one of the published data sets from the VLDS catalog. Simply incorporating a small piece of the data in an application (i.e., as an afterthought) was insufficient. To be competitive, developers must have created new ways for stakeholders and/or the public at large to analyze, understand, and make informed decisions based on the data. Mashups with other data sources were not only allowed, but encouraged. Cash prizes were offered for the winners of the various competitions as incentives to participate.

1.2 VLDS Data Sets

The following nine data sets are available online for researchers and developers. Each data set is a collection of CSV files, with one file for each year. In January 2013, the files comprised about 4.5 GB of data. The data sets are well documented, and a dictionary of each attribute is available on the VDOE website.⁵ Each row of data is uniquely identi-

¹For more information on the Virginia Longitudinal Data System, visit http://www.doe.virginia.gov/info_management/longitudinal_data_system and vlds.virginia.gov.

²CIT is a state-chartered non-profit organization with a mission to accelerate technology development in the Commonwealth of Virginia. Recognizing a need in the marketplace, CIT Connect was established to help federal, state, and corporate consumers of technology to identify and assimilate innovation.

³In addition to the general public, some of the communities targeted were computer programmers, software developers and designers, Web designers, researchers, entrepreneurs, and educators.

⁴The hackathons were 24-hour events in which computer programmers, software designers, graphic designers and others formed teams to develop usable software applications.

⁵See http://www.doe.virginia.gov/statistics_reports/research_data/

fied using the following values: school year, division number, school number, grade level, and demographic group (which includes race, gender, disabled status, English proficiency, and disadvantaged status).

- **Fall Membership:** Number of students enrolled in Virginia public schools on Sep 30th.
- **Dec 1st Count:** Number of students with disabilities receiving special education.
- **Test Data:** Average scores and pass/fail rates for English, History, Math, and Science.
- **CTE Completers:** How many students completed Career and Technical Education course sequences.
- **HS Graduates:** Number of high school graduates and completers of similar diplomas.
- **Annual Dropouts:** Summer and term dropouts in grades 7–12 (who did not return by Oct 1st).
- **Cohort Rates:** On-time graduation rates for students entering 9th grade together.
- **Postsecondary Enrollment:** High school graduates enrolled in postsecondary institutions nationwide.
- **Postsecondary Achievement:** High school graduates enrolled in public higher education in Virginia who earned one year of college credit within two years.

What makes these data sets particularly interesting for use in a university-level database course is that the underlying information is both *aggregated* and *suppressed*. As stated on the download site, "The datasets are compiled using all the possible combinations of all the demographics about students, so each row within the dataset contains a rate or count in addition to the demographics used to arrive at the rate or count." And further, "Within each dataset, rows were withheld if deemed that the number of students in the group could lead to the identification of a single student. In most cases, student groups of nine or fewer are suppressed."

1.3 About the Course

Given the richness of the VLDS data and the availability of materials from prior competitions, we decided to adopt Apps4VA as a semester-long project. Our course is an introduction to database systems that uses the textbook from Ullman and Widom [12]. One expected learning outcome of the course is that students will design and implement a substantial three-tier application. At this point in our CS department's curriculum, students have completed the introductory programming sequence and courses in data structures and software engineering. The database course is an ideal opportunity for students to synthesize what they have learned in these prerequisites.

A major benefit of this project-based learning approach is that students have a meaningful running example throughout the course. When we first looked at how data can be represented (e.g., entity-relationship model, relational model, semi-structured model), we illustrated the pros and cons of each model using VLDS data. Later in the course, students learned how to write relational algebra and SQL statements using VLDS data. In addition, we were able to motivate

functional dependency theory using VLDS, especially since the data had been joined and aggregated prior to release. As a result, students not only learned the content of the course, they also became familiar with nuances in the data that, ultimately, would direct their semester projects.

2. PROJECT DELIVERABLES

Students worked on their Apps4VA projects throughout the semester, submitting work about every two weeks. In this section, we briefly describe the nature of the assigned deliverables and how they were assessed, with the intent that readers will be able to adopt similar projects in their own courses. Additional information about each project is available on our website: <http://apps4va.cs.jmu.edu/>

In order to stimulate creativity and to help students to imagine the types of applications that could be developed, we discussed the submission of a finalist of the Apps4VA Open Competition.⁶ This developer generated a KML file suitable for Google Earth that enabled users to navigate the VLDS data set in new and interesting ways. In general, students could be allowed to develop a wide variety of solutions ranging from traditional desktop software to mobile apps and web services. But to level the playing field and to simplify the administration of the projects, we required our students to develop a web application using Servlets and JSPs. (Java is the primary language of our introductory course sequence.)

The specific language and database technology being used are not as important as being consistent within the course. In previous years (before we adopted Apps4VA), we allowed students to make their own technology choices with the understanding that the instructor would not be able to provide support beyond examples shown in class. But doing so led to problems with one or two students being enthusiastic about their chosen framework (e.g., Ruby on Rails) while the rest of their team became frustrated and unable to contribute to the project.

2.1 Written Proposal

As with many semester-long projects, we began by having students write a short proposal (750 to 1000 words). We require this assignment as early as possible, typically at the beginning of the second week. The early assignment, in turn, requires the students to select their teammates during the first week of class. To help students understand the project as a whole and to help them to decide on the type of applications they would build, we directed them to review the Apps4VA website. This website has several resources⁷ that we found to be useful.

- *Educational Data Primer*: Describes what data is collected, why it is important, and how it can be used to make improvements to the K-12 educational system.
- *Idea Bank and Blog*: Highlights the needs, requests, and suggestions from teachers, administrators, and policy makers, as well as the community at large.
- *Data Sets*: Provides detailed instructions on accessing the required data from VLDS, as well as optional data from other governmental agencies (e.g., census records, health statistics).

⁶See <http://www.starsman.com/VEGAS.html>

⁷See <http://www.apps4va.org/resources.html>

Students were instructed to spend at least two hours reviewing the website's resource information before starting their actual write-ups. We found that the most important aspect of this assignment was for students to identify a "specific K-12 education research problem" to be addressed through their projects. Having students focus on education research not only made the project more focused, it also prevented students from adding unnecessary features for the sake of turning in copious amounts of code at the end of the semester.

We spent nearly an entire class period for students to review proposal submissions from other teams. Each team received a folder containing three proposals and three copies of a rubric that was available to them when the assignment was posted. They were to spend ten minutes reading and marking up each proposal as a team, and then filling out the rubric to determine its grade. We then collected the rubrics and set up a rotation for each team to meet with the authors they reviewed and give to constructive feedback for improving the proposal.

This peer assessment encourages collaboration, facilitates the exchange of ideas, communicates expectations, and provides an active learning environment [5]. Students were encouraged to revise and resubmit their work by the next class period. As a result, the research questions and project ideas were clearly specified, and the writing quality was much better when the instructor reviewed the proposals in detail.

2.2 Schema Design

The next phase of the project required students (1) to create an initial set of tables to support their applications and (2) to write scripts designed to extract, transform, and load VLDS data into their database. Most of our students have little, if any, prior experience designing databases. At this point of the semester, they are familiar with relational algebra and SQL, but have not yet studied functional dependencies and normalization. We found that having students work through the design process early motivates theoretical topics later in the course.

We also introduced students to Subversion and required them to maintain all project-related files in a repository. This practice not only helped them to develop a useful skill, it made it easier for them to make changes to the database design throughout the semester [9]. During their subsequent projects and homework assignments, students learned how to improve their work, identify what was missing, and/or develop new ideas. Since their entire databases were created using scripts, they were free to experiment with incremental changes without the risk of losing their work.

To assist the instructor in assessing their designs, we required students to create a relational database diagram that indicated the number of rows imported into each table. Most of the major mistakes that students made in writing the queries that transformed the VLDS data became apparent by looking at the row counts. The tasks of designing and importing real data reinforced one other and, in turn, helped students to understand the nature of the relational model.

2.3 Example Queries

For this project, students had complete flexibility in the direction they took. The basic requirements were (1) to revise their schema based on instructor feedback and (2) to design and implement ten queries their application would

use. As in the previous assignment, the two steps reinforced one other. As students began to design queries, they discovered false assumptions they made about the underlying data sets. They also identified aspects of their designs that made writing some queries more difficult than necessary. Once they completed this assignment, they were ready to begin analyzing the data and answering the educational research problems they originally proposed.

For those interested in adopting this type of semester project, some deliverables should, ideally, include a formative community-based assessment. So that students might have the benefit of a real-world perspective, we invited three representatives from industry and government to meet with each group of students during class time. These visitors were experts in the field of data analytics for K-12 education, and their involvement generated a high level of engagement among the students. Each group had the opportunity to explain their research problems and demonstrate results from some of their queries. The feedback students received greatly influenced the direction of their projects by identifying additional resources and refining their educational research questions.

2.4 Detailed Design

Now that students had proposed, designed, and “sold” their projects, they were ready to develop their final applications. The deliverable for this assignment was a detailed list of functional requirements and a table that outlines which JSPs will implement which features. In addition, students were to fully implement at least one web form. Since most students had little or no experience in web development, the goal of this assignment was to give them a big picture of three-tier architecture and to clarify the technical steps of setting up and deploying their projects. In addition, students spent time learning any third party tools or APIs they planned to use, such as Google Charts and Google Maps.

The feedback that students received from this project was primarily summative. Most, if not all, student groups were able to get something up and running by this point. Some feedback about their choices of requirements was necessary in order for all to agree on a final deliverable, but that often happened during class time while we discussed each group’s progress on their design.

2.5 Initial Prototype

We deliberately made this deliverable the most important deadline of the semester. Rather than giving students until the last week of class to complete their projects, we evaluated their prototype systems two weeks before their final presentations. The only instruction they were given was that they had to complete the requirements they submitted in the previous assignment. Students were provided a rubric that specified the review criteria for three aspects of their projects: functionality, performance, and code quality. The instructor evaluated each project during class time so that students could conduct an interactive demonstration of their work. Specific follow-up requirements were given for students to complete by the end of the semester.

2.6 Final Presentation

Finally, students presented their work to the class and to the industry and government mentors described previously. Their presentations focused on the educational re-

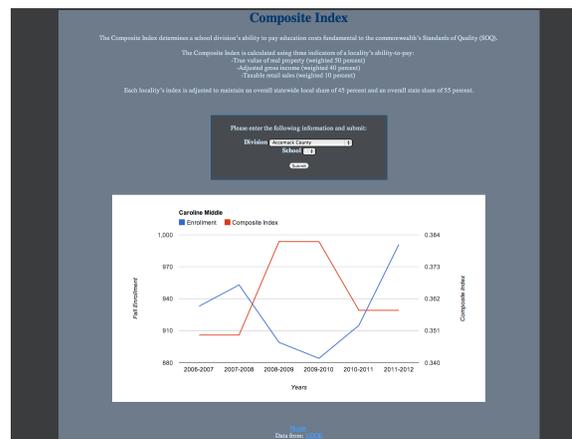


Figure 1: Screenshot of Graduation Rates vs Funding Levels

search problems addressed by their applications, their applications’ main features, the primary users and beneficiaries of their work, what improvements to education they might expect to result from their applications, and any future work they have identified. As with the third deliverable, we incorporated a community-based assessment into this final presentation. The industry and government collaborators participated as review panelists. They congratulated students on their hard work and offered real-world critiques and insights on their applications and the results they generated.

3. EXAMPLE APPLICATIONS

We now present three web-based applications that students developed during the course. These concrete examples illustrate the types of projects students can pursue using the VLDS data. The particular applications we have selected are representative of those developed by all student teams and provide context for the outcomes described in the next section of the paper.

3.1 Graduation Rates vs Funding Levels

One team of students was particularly interested in helping educators and administrators prepare more effectively for future student enrollment. They also were concerned about how tax dollars would be spent: “Our application enables policy makers and educators to view information on different student groups by referencing divisions, schools, test scores, and graduation rates in Virginia in order to distribute funding to the areas that are most in need.” Figure 1 shows how the application compares a school division’s enrollment count (lower line) with the community’s ability to pay education costs (upper line). This “composite index” is based on property values, adjusted gross income, and taxable retail sales. In this example, enrollment suddenly increased while the composite index decreased.

3.2 Education Programs vs Workforce Needs

Another team asked two questions that often resonate with parents of high school students: “(1) Are students being prepared for real-world jobs? (2) What jobs are in demand in their area of the state?” The students explained, “Our application allows policy makers and superintendents of school

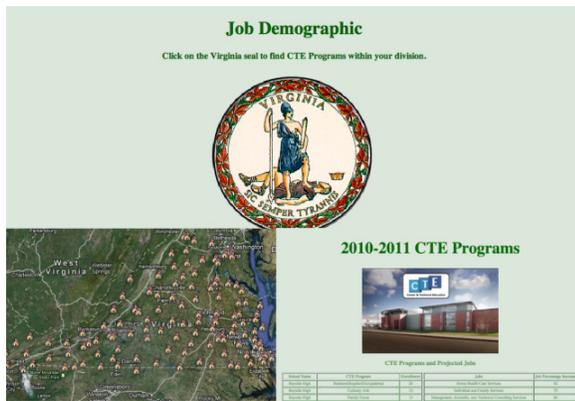


Figure 2: Screenshot of Education Programs vs Workforce Needs

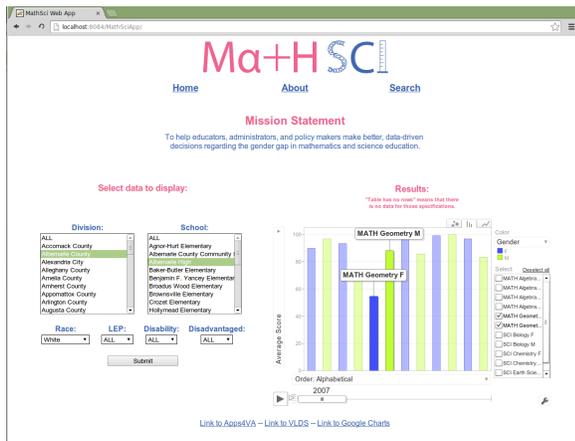


Figure 3: Screenshot of Gender in Math and Science Education

divisions to look at the industry projections for different careers and match them to the programs they offer at their career and technical education schools. The app is based on the CTE Program Graduates and Completers VLDS 2010–2011 data, as well as Labor Force and Industry Projections from the Virginia Workforce Connection. These individuals will utilize our app to determine which CTE programs they need to offer based on the top industry projections.”

As illustrated in this example, most student groups developed applications for administrators, policy makers, and teachers. However, these types of users already have access to sophisticated tools and reports tied directly to VLDS. One lesson we learned during the mid-semester and final presentations is there is a substantial need (and an emerging market) for tools that focus on local communities. If parents, students, and local leaders understand what the data is saying about their schools, they will be in a better position to ask for meaningful changes in curricula and programs.

3.3 Gender in Math and Science Education

Our final example project examines the gender gap in mathematics and science test performance across grade levels. These students leveraged motion charts to “enable policy makers and administrators to view data regarding mathematics and science education among different genders and

identifying factors.” Figure 3 shows the relative test scores between males and females at a specific school (selected on the left). Each pair of bars represents the math and science exams offered in each grade level. By clicking the play button or dragging the slider, the user can see how scores increase and decrease over time. This visualization makes it relatively easy for someone unfamiliar with VLDS to understand the types of data available and to use it in a meaningful way.

4. STUDENT OUTCOMES

In comparison to previous offerings of the course, students involved with Apps4VA were particularly engaged in their semester-long projects. Although we did not conduct a rigorous study to measure student attitudes quantitatively, we do have some data points that indicate our adoption of Apps4VA was successful. For one, the number of students who withdrew or failed the course greatly decreased as shown in Table 1. (Note: Students who withdraw from the course before the mid-semester deadline receive a W grade on their transcript. Since W grades are not counted in GPA calculations, this action is recommended to students who are likely to receive an F grade at the end of the semester.)

Year	D	W	F	Rate
2012	4	6	4	23%
2013	7	1	1	15%

Table 1: Comparison of D, W, and F grades (same course, same instructor, 60 students each year)

The course project in 2012 introduced students to computer science research literature by requiring them to build a web application featuring citation data from DBLP.⁸ We required the same deliverables as those described in this paper, and we used DBLP as a running example throughout the course. However, we did not have a community-based assessment (i.e., external collaborators reviewing the students’ work) and the resulting projects simply were not intended to impact society at large. These factors may have contributed to a higher level of cohesion and commitment among this year’s group members.

We did not see a significant difference in overall student performance in the course, in terms of their final exams or individual project scores. However, the overall rigor and expectations of the course were much higher in 2013. Both the quality of the final presentations and student feedback on course evaluations showed vast improvement from the previous offerings. Based on these assessments, students appeared to learn more content and learn it more efficiently than in prior years.

After the course concluded, one of the organizers of the Apps4VA competitions who helped to evaluate this year’s final presentations observed the following: “While one of the goals of the program was to discover creative uses for the data, it turned out that discovering creative stakeholders and participants was a surprisingly welcome outcome. The apps developed by these student developers rivaled and, in some cases, eclipsed those designed by professional developers.” A number of students indicated via email and in

⁸See <http://dblp.uni-trier.de/>

private conversations that Apps4VA gave them something concrete to discuss in their job interviews.

5. RELATED WORK

This work documents effective practices for the teaching of database systems and data analytics. In recent years, a number of experience papers have proposed complementary techniques for enhancing database courses. [4] demonstrated promising results by incorporating a “Contributing Student Pedagogy” where groups of students become responsible for the delivery of course content. The paper reports on two iterations using this teaching method and shows that having multiple deliverables helps students to produce higher quality presentations. [7] explored the use of “Just-in-Time Teaching” in a database course, where students completed online warm-up assignments prior to each class. Warm-ups make it possible to address students’ specific needs, and the paper shows how this practice improves exam scores and final grades. [3] introduced a suite of exercises using “Language INtegrated Query” (LINQ) in the database curriculum. The paper showed how LINQ not only can meet learning objectives of typical database assignments, but also support program outcomes by exposing students to functional programming.

Two important aspects of our work include community-based assessments and collaborative student projects. As shown in [2], these “developmental interaction experiences” lead to more career development, business knowledge, and psychosocial support among students. Our work is also related to the growing body of literature that studies No Child Left Behind and other recent legislation. The references in the Introduction of this paper provide a good starting point. As computer science students begin to wrestle with these complex social issues within the context of their major courses, they make connections to the human dimension that results in significant learning [5].

6. CONCLUSION

We have described our experiences in integrating K-12 education research into an introductory database course. Our hope is that other instructors will achieve similar results and benefits by implementing these projects in their courses. The data sets provided by Apps4VA and VLDS are rich and meaningful, well documented, and freely available. We encourage readers of this paper to form partnerships with departments of education and local governments to make similar data sets available from their respective states and countries. This type of application development has the potential to impact students and local communities in many positive ways. In our experience, students were motivated to develop successful, high-quality projects because the data was about them and the future of their educational system.

7. ACKNOWLEDGMENTS

We thank the students of teams *Infinity*, *JobDemo*, and *MathSci* for allowing us to write about their projects, use their screenshots, and quote their presentations.

8. REFERENCES

- [1] K. Anthes. *School and District Leadership*. No Child Left Behind Policy Brief. Education Commission of the States, Denver, CO, USA, 2002.
- [2] C. D’Abate. Developmental Interactions for Business Students: Do They Make a Difference? *Journal of Leadership and Organizational Studies*, 17(2):143–155, 2010.
- [3] S. W. Dietrich and M. Chaudhari. LINQ ROX!: Integrating LINQ into the Database Curriculum. In *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education*, SIGCSE ’11, pages 293–298, 2011.
- [4] H. A. Etlinger. Adding a Contributing Student Pedagogy Component to an Introductory Database Course. In *Proceedings of the 44th ACM Technical Symposium on Computer Science Education*, SIGCSE ’13, pages 299–304, 2013.
- [5] L. D. Fink. *Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses*. Jossey-Bass, San Francisco, CA, USA, 2003.
- [6] J. C. Kabaker. 20 States Receive State Longitudinal Data Systems Grants. New America Foundation, Washington, DC, USA, May 2010.
- [7] A. Martinez. Using JITT in a Database Course. In *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education*, SIGCSE ’12, pages 367–372, 2012.
- [8] R. Reback, J. Rockoff, and H. L. Schwartz. *Under Pressure: Job Security, Resource Allocation, and Productivity in Schools Under NCLB*. NBER Working Paper 16745. National Bureau of Economic Research, Cambridge, MA, USA, 2011.
- [9] D. Rocco and W. Lloyd. Distributed Version Control in the Classroom. In *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education*, SIGCSE ’11, pages 637–642, 2011.
- [10] A. P. Scandrett. *Principals’ Perspectives of the No Child Left Behind Act of 2001 as They Relate to Personnel in a Rural School System*. PhD thesis, Mercer University, 2010.
- [11] J. Supovitz, E. Foley, and J. Mishook. In Search of Leading Indicators in Education. *Education Policy Analysis Archives*, 20(19), 2012.
- [12] J. D. Ullman and J. Widom. *A First Course in Database Systems*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2007.