

Can Multimedia Tools Promote Big Data Learning and Knowledge in a Diverse Undergraduate Student Population?

Sinjini Mitra¹ and Archana J. McEligot²

¹ *Department of Information Systems & Decision Sciences, California State University, Fullerton*

² *Department of Health Sciences, California State University, Fullerton*

Abstract

Background and Purpose: Multimedia tools are an integral part of teaching and learning in today's technology-driven world. The present study explored the role of a newly-developed video introducing the emerging field of big data to a diverse undergraduate student population. Particularly, we investigated whether introduction of a multimedia tool would influence students' self-perceived knowledge related to various big data concepts and future interest in pursuing the field, and what factors influence these.

Methods: Students ($n = 331$) completed a survey on-line after viewing the video, consisting of Likert-type and quantitative questions about students' learning experience, future interest in big data, and background. The dataset was analyzed via ANOVA and multiple linear regression methods.

Results: Gender, major, and intended degree were significantly associated with students' learning experience and future interest in big data. Moreover, students who had no prior exposure to big data reported a better learning experience, although they also reported less likelihood to pursue it in the future. **Conclusion:** Multimedia tools may serve as an effective learning tool in introducing and creating interest in a diverse group of students related to introductory big data science concepts. Both similarities and differences were observed regarding such behaviors among different student sub-groups.

© 2018 Californian Journal of Health Promotion. All rights reserved.

Keywords: big data, multimedia, video, learning experience, career pathways

Introduction

Big data refers to data sets that are so large and complex that they cannot be processed and analyzed by traditional applications and software (Holmes, 2017). As technology continues to evolve and grow at a rapid pace today, driven mainly by the development of smart devices, sensors, and cloud computing, more and more data collected from various sources in different industries including healthcare, education, and finance, are becoming accessible to researchers in an unprecedented way. The huge social and academic impact of such developments caused a worldwide buzz for "big data" as new technologies slowly started to emerge that are able to store, process and analyze such data. The concept of big data, commonly characterized by volume (amount of data), variety (diverse and complex nature of variables included in the data such as text, images, videos), velocity (the speed or rate at which the data become available), and veracity (how much noise and uncertainty are

there in the data), goes far beyond traditional data types, as well as statistical analyses using common descriptive and inferential methods. According to an article in Forbes magazine (Marr, 2015), by the end of 2020, about 1.7 megabytes of new information will be created every second for every human being on the planet. This is equivalent to about 11-12 standard digital photo files, thus showing the enormity of the data explosion happening in today's society.

Driven by the need to improve healthcare delivery and cost in efforts to improve health, these massive amounts of data hold the promise of supporting a wide range of medical and healthcare needs/areas related to big data science and/or big data biomedicine, such as, clinical decision support, disease surveillance, neuroimaging/brain health, patient data in electronic health records (EHRs), insurance

claims data, and population health management (Demboosky, 2012; Huang et al, 2015). Given this widespread use and need, it is of utmost importance to train students of today in this fast-evolving area of the applications of big data in biomedicine. Although concepts of statistical and analytical methods are included in the curriculum of most college degree programs in the sciences today, not many students have a good understanding of what big data or data science is or how different they are from traditional data that are typically utilized in their courses and projects and what the underlying issues and challenges are in handling and analyzing such data (Frost & Sullivan, 2012).

Multimedia is defined as content that uses a combination of audio and video representations such as text, audio, images, animations, video, and interactive media (Liaskos & Diomidus, 2002). Multimedia tools play a big role in today's higher education as technology is increasingly being integrated in college courses in the form of primarily videos and interactive activities like games or animations in order to foster more active learning. According to a report published in Ed Tech (Ed Tech, 2016; Ilan & Oruc, 2016), some benefits of multimedia learning include deeper understanding, improved problem-solving skills, increased positive emotions, and access to a vast variety of information. It provides rich opportunities to augment current educational techniques and creates content that are attractive and engaging for students. Multimedia is used on a daily basis by the current generation of learners, like millennials, today, thus making it indispensable and in keeping with the progression of the world today (Tang, 2014). Moreover, psychologists have shown that information obtained through visualization is retained 83% of the time compared to only 11% retained through hearing and 5% through other senses (Yang, 2010). Similar experiments have also established that people can remember 50% of content presented via audio and images combined (as in a video) whereas only 10% is remembered via reading and 20-30% via only audio or video. Chen and Xie (2011) states that the use of videos and interactive media allow students to apply the

knowledge learned and stimulates creativity, enthusiasm and involvement.

Therefore, we explored the role of a multimedia tool, namely a video that incorporates audio, images and text, in educating undergraduate students on introductory big data science concepts, including defining big data, identifying big data, challenges and applications of big data, as well as potential opportunities in big data/data science. Given that big data is a newly emerging field and its significance in today's world where data are generated in massive amounts, there is a need to educate students about the topic and its relevance and applications in different fields. Therefore, we sought to test whether the implementation of a short video during class increases perceived data science knowledge for our diverse undergraduate students and other sub-groups.

Further, other sub-aims included investigating whether familiarity with big data prior to watching the video had any effect on students' learning experience with the video, as well as on their subsequent interest in exploring this field in the future, whether to just learn more using additional resources or even pursue career options.

Methods

Setting

A video was developed and shown in introductory statistics, mathematics and epidemiology courses. After watching the video, students in those classes filled out an online survey questionnaire about their perceived learning/educational experience related to the video. The questionnaire consisted of both Likert-type and quantitative questions regarding the students' learning experience, future interest in pursuing the topic, as well as questions about their academic and demographic background and prior exposure to the topic. The study protocol review was conducted and approved by the Internal Review Board (IRB) of the California State University, Fullerton (HSR# 16-0221).

Participants and Inclusion/Exclusion Criteria

Participants in the study included undergraduate students enrolled in introductory statistics,

mathematics and epidemiology courses from across different departments such as Mathematics, Biology, Geology, Communications, Business, Health Science, and Computer Science. Although video viewing in these courses may have constituted a convenience sample (not probability-based), it served the purpose of our study as our aim was to educate students in these disciplines about big data, because they are and will be most likely exposed to such data in their respective fields. Students were from diverse demographic backgrounds, and were at least 18 years old. No willing student/participant was excluded from viewing the video and providing responses.

Approximately 800 students were enrolled in the classes that were invited to participate in the study. Of these students, 331 responses were received, out of which 21 participants were excluded because they did not answer at least half of the questions on the survey. This produced an analytic sample size of $n = 310$ for our study. The participation rate was not calculated because it was unknown to the researchers as to how many instructors actually showed the video in their classes.

A majority were female students (63.6%). With regards to ethnicity, Hispanics (32.6%) and Asians (30.9%) made up the majority of the survey respondents, while 22.4% were Caucasians. The mean age of the students who participated in the survey was 22.13 years (with a standard deviation of 3.6 years). Thus, this sample was a good representation of the overall undergraduate student population in the university which is dominated by under-represented minority groups (Office of Institutional Research and Analytical Studies, CSUF).

As for academic background, nearly 54% were Juniors, 25% Seniors, and 20% Sophomores. Moreover, 36% of student participants wanted to pursue a Master's degree, and 27% aspired to earn a doctoral or other advanced professional degree. As to the students' intended or declared majors, majority of them were from Health Science (48%), followed by Biological Science (12.3%) and Mathematics (6%). Moreover, there

were students from other disciplines such as, Computer Science, Business, Communications, and Geology.

Multimedia Tool Development and Content

The multimedia production predominantly focused on developing the video: *Big Data Science: An Introductory Tutorial*. The making of the video involved a comprehensive process, beginning with collaborative discussions between CSUF faculty, students and staff, and an external video production team. Through an iterative process, the production team and the program director outlined the final video structure, highlighted important themes, articulated specific educational goals, crafted an interview questionnaire for the faculty/experts who participated in the video, began to create a production schedule, and completed interviews with faculty/experts in their respective big data science fields about their own experiences and specific issues and challenges in working with such data in their fields. A video treatment, taking into consideration the feedback of all stakeholders was created which was reviewed and revised for any final changes. The final educational product was thus the culmination of a creative and analytical process designed to produce a film/video that would educate and engage students in introductory big data concepts. The video is posted on YouTube and can be accessed via the following link: https://www.youtube.com/watch?v=25Z-iALT_KM&t=161s.

The content included the following: *Part 1*: What is Big Data, including discussion on the advent of Big Data, defining Big Data. *Part 2*: Big Data Sources: General Sources – Open Sources, and a specific focus on genomics and epidemiologic data types. *Part 3*: Introduction to Data Synthesis, including data management, mining and visualization, and summarization. *Part 4*: Solutions using Big Data. *Part 5*: Big Data Career Pathways. The entire video was 34 minutes and 53 seconds long.

Data Collection: Survey Instrument

A survey instrument was developed by the research team in consultation with an external evaluation firm primarily to assess perceived

student learning related to introductory big data science concepts after viewing the video. Specifically, the survey queried on perceived understanding/learning of big data, underlying issues and challenges and strategies to handle big data, and also to investigate their interest in pursuing the field further in the future as a result of watching the video. These form the two main constructs of our study.

The construct for measuring students' perceived learning experience via the video consisted of 10 Likert-scale type questions (6-point scale: 6: *Strongly agree*, 5: *Agree*, 4: *Slightly agree*, 3: *Slightly disagree*, 2: *Disagree*, 1: *Strongly disagree*) to assess students' self-perception about the level of learning regarding various aspects of big data, such as:

- i. understanding the importance of big data in health science today;
- ii. ability to define big data;
- iii. understanding what the different elements of big data are – the 4 V's;
- iv. understanding the challenges posed by big data environments, and some potential ways to address them;
- v. understanding the various sources of big data in different fields, such as, health, biomedicine, mathematics and business;
- vi. understanding the applications of big data in health and biomedical sciences;
- vii. understanding research on big data in health and biomedical sciences;
- viii. understanding how to manage big data in the context of health and biomedical sciences;
- ix. understanding some statistical tools necessary to visualize, summarize and analyze big data in the context of health and biomedical sciences; and
- x. overall learning experience.

The survey also queried students about their level of familiarity with big data before watching the video using a 4-point Likert scale (1: *Very familiar*, 2: *Somewhat familiar*, 3: *Not very familiar*, 4: *Not at all familiar*). Moreover,

it sought information about the students' demographic and academic background like age, gender, ethnicity/race, major, and class status. Finally, the survey concluded by asking students regarding their interest in exploring additional materials (e.g., courses and tutorials) and career pathways in the field of big data applications in health and biomedical sciences. This construct was measured with the help of a similar 6-point Likert scale (6: *Strongly agree*, 5: *Agree*, 4: *Slightly agree*, 3: *Slightly disagree*, 2: *Disagree*, 1: *Strongly disagree*) that was used for the learning experience construct.

Data Analyses

Prior to our data analyses, we ran Cronbach's alpha (Cronbach, 1952) to test the reliability of the scales used to measure students' learning experience and their interest in pursuing the field of big data further in the future, both of which consisted of multiple item sets. The values obtained were respectively 0.936 and 0.871, hence demonstrating a high level of internal consistency and reliability.

Apart from initial descriptive statistics to summarize the demographic and academic composition of our sample, we used statistical inference methods: analysis of variance (ANOVA) and multiple linear regression to test the hypotheses. To study students' learning experience about big data after watching the video, we created two outcome variables: (1) overall learning experience rating, and (2) average rating for the first nine items on the Likert scale. The rationale for using the combined measure instead of nine separate outcome variables was that our preliminary analyses revealed that there was not much difference in the ratings for the nine individual questions with a significant amount of inter-item correlation.

ANOVA was performed to investigate if differences existed in the outcome variables representing students' learning experience by background academic and demographic variables namely, gender, ethnicity, major, class level, and intended degree (all categorical variables). ANOVA was also performed to examine the differences in prior familiarity with

big data and interest in pursuing the big data field (both for exploring additional resources and career pathways), also by academic and demographic variables.

The overall learning experience construct was chosen as the dependent or outcome variable for the multiple regression model to study the effect of prior familiarity with the topic, along with adjusting for the independent variables, including age (continuous variable), gender, ethnicity (categories used: Hispanics, Whites, Asians, and “other” as the reference category), major (categories used: Mathematics, Health science, Biology, Chemistry/Biochemistry, and “other” as the reference category), class level (categories used: sophomore, junior, senior, and freshman as the reference category), and intended degree (categories used: Master’s, PhD and other professional, and Bachelor’s as the reference category).

To study the relationship between prior big data familiarity and interest in pursuing the big data field in the future, the dependent or the outcome variable was formed by the mean of the responses to two survey questions – (i) level of interest in exploring additional materials and resources and (ii) level of interest in exploring career pathways in the area. Prior familiarity was included as the dependent variable in the model while adjusting for the same covariates with same categories (including reference) as those used in the case of the studying students’ perceived learning experience.

Learning experience from the video can also be a driver of students’ interest in future pursuance of the field; however, both variables denoting learning experience and prior familiarity were not included in a regression model together because of the significant association between the two as evident from the correlation analysis ($r = 0.577$, with $p < 0.05^*$), introducing multicollinearity in the model.

All the analyses were performed using the statistical software package IBM-SPSS Version 24 on a Macintosh computer (IBM SPSS

website). For all the analyses, 0.05 was chosen as the criterion for statistical significance based on p-values, and findings with p-values between .051 and 0.10 were discussed as demonstrating borderline significant patterns as well.

Results

Learning Experience

The mean overall rating for students’ perceived learning experience was 5.74 (out of a 6-point Likert scale) with a standard deviation of 1.19 ($n = 290$), showing that students generally agreed that watching the video was a valuable experience in relation to big data science. The average score for the ten individual items included in the survey instrument was 5.50 (also on a 6-point Likert scale) with a standard deviation of 1.005, also showing a positive learning experience of students about all the aspects of big data covered in the video. Further, the low standard deviation indicated consistency of the students’ ratings regarding understanding of different aspects of big data and its applications.

Table 1 shows the primary descriptive statistics (frequencies, mean \pm standard deviation) of the two outcome variables representing students’ perceived learning experience (overall rating and average of the ratings for the 9 individual items) by gender, ethnicity, major, and class level. Main findings related to the different student sub-groups can be summarized as follows: (i) Females reported a statistically significantly ($p = 0.052$) greater overall perceived learning experience than males. Students intending to pursue a PhD or other higher professional degree reported borderline significant higher mean ratings for both outcome variables mentioned above (p values of 0.083 and 0.077 respectively), the lowest mean ratings being for those pursuing a Bachelor’s degree alone. No significant differences were observed with respect to ethnicity, class level, and major. This showed that the effectiveness of the video was uniform across student sub-groups defined by these background factors.

Table 1

Descriptive Statistics for Video Ratings by Students' Demographic and Academic Background Variables					
	N (%)	Overall learning experience		Average of ratings over 9 items	
		(Mean ± SD)	p	(Mean ± SD)	p
Overall		5.74 ± 1.19		5.50 ± 1.01	
Gender	305		0.052*		0.242
Male	111 (36.4%)	(5.56 ± 1.31)		(5.40 ± 1.04)	
Female	194 (63.6%)	(5.84 ± 1.11)		(5.55 ± 0.98)	
Ethnicity	304		0.992		0.649
American Indian/Alaskan native	4 (1.3%)	(5.67 ± 1.15)		(5.56 ± 0.89)	
Asian/Asian American	94 (30.9%)	(5.71 ± 1.22)		(5.39 ± 0.99)	
Black/African American	6 (2%)	(5.60 ± 0.55)		(5.20 ± 0.84)	
Hispanic	99 (32.6%)	(5.75 ± 1.29)		(5.46 ± 1.14)	
Native Hawaiian/Pacific Islander	5 (1.6%)	(5.40 ± 0.55)		(5.33 ± 0.42)	
White	68 (22.4%)	(5.73 ± 1.19)		(5.63 ± 0.91)	
Multiethnic	16 (5.3%)	(5.80 ± 0.78)		(5.75 ± 0.72)	
Other	12 (3.9%)	(6.00 ± 1.13)		(5.82 ± 0.97)	
Major	302		0.388		0.443
Biological Sciences	37 (12.3%)	(5.78 ± 0.96)		(5.65 ± 0.77)	
Mathematics	17 (5.6%)	(5.82 ± 1.51)		(5.42 ± 1.35)	
Health Sciences	145 (48%)	(5.80 ± 1.09)		(5.56 ± 0.93)	
Psychology	4 (1.3%)	(6.25 ± 0.50)		(5.75 ± 0.60)	
Biochemistry	14 (4.6%)	(6.21 ± 0.58)		(5.83 ± 0.41)	
Chemistry	2 (0.7%)	(6.00 ± 0.00)		(5.39 ± 0.71)	
Other	83 (27.5%)	(5.53 ± 1.35)		(5.32 ± 1.10)	
Class level	293		0.632		0.269
Freshman	2 (0.7%)	(5.50 ± 0.71)		(5.89 ± 1.26)	
Sophomore	59 (20.1%)	(5.87 ± 1.06)		(5.54 ± 0.89)	
Junior	156 (53.2%)	(5.66 ± 1.28)		(5.40 ± 1.08)	
Senior	76 (25.9%)	(5.80 ± 1.12)		(5.66 ± 0.94)	
Intended Degree	284		0.083*		0.077*
Bachelor's	100 (35.2%)	(5.50 ± 1.36)		(5.41 ± 1.15)	
Master's	104 (36.6%)	(5.79 ± 1.14)		(5.43 ± 0.99)	
Ph.D., Ed.D., MD, other prof.	80 (28.2%)	(5.95 ± 1.02)		(5.73 ± 0.82)	

Note. The p-values reported are from the ANOVA tests studying the difference among the different groups. P-values: *<0.10, **<0.05, ***<0.0001.

Association between Learning Experience and Prior familiarity with Big Data

The overall mean familiarity with big data was 2.74 (on a 4-point Likert scale) with a standard deviation of 0.836 (*n* = 304). Moreover, 22.4% of students were totally unfamiliar with big data

(that is, they had never even heard the term prior to watching the video), 37.5% were not very familiar (that is, had just heard the term “big data” without even knowing what it meant). Table 2 shows the mean familiarity ratings (and associated standard deviations) across the different factors.

Table 2

Descriptive Statistics for Familiarity with Big Data Prior to Watching the Video by Students' Demographic and Academic Background Variables

	N (%)	Familiarity with Big data before video	
		Mean ± SD	p
Overall		2.74 ± 0.84	
Gender	305		0.047**
Male	111 (36.4%)	2.66 ± 0.78	
Female	194 (63.6%)	2.85 ± 0.86	
Ethnicity	304		0.481
American Indian/Alaskan native	4 (1.3%)	3.25 ± 0.96	
Asian or Asian American	94 (30.9%)	2.77 ± 0.91	
Black or African American	6 (2%)	3.17 ± 0.98	
Hispanic	99 (32.6%)	2.78 ± 0.80	
Native Hawaiian/Pacific Islander	5 (1.6%)	2.60 ± 0.55	
White	68 (22.4%)	2.75 ± 0.76	
Multiethnic	16 (5.3%)	2.56 ± 0.81	
Other	12 (3.9%)	3.17 ± 0.94	
Major	302		0.041**
Biological Sciences	37 (12.3%)	2.59 ± 0.90	
Mathematics	17 (5.6%)	2.65 ± 0.61	
Health Sciences	145 (48%)	2.83 ± 0.85	
Psychology	4 (1.3%)	3.50 ± 0.58	
Biochemistry	14 (4.6%)	2.71 ± 0.61	
Chemistry	2 (0.7%)	2.50 ± 0.71	
Other	83 (27.5%)	2.77 ± 0.86	
Age		correlation: 0.084	0.180

Notes. Since age is reported as a continuous variable (no categories), a correlation coefficient was computed in place of ANOVA to study its association with Big Data familiarity. The p-values reported are from the ANOVA tests studying the difference among the different groups. * <0.10 , ** <0.05 , *** <0.0001 .

Male students were significantly ($p = 0.047$) more familiar with big data prior to watching the big data video. Similarly, familiarity with big

data varied significantly across students' major field of study ($p = 0.041$). Psychology students were the least familiar with this emerging field while chemistry and biology students were the most familiar. A negative correlation with age (continuous variable) indicated that younger students were less familiar with the field prior to watching the video in their classes.

Table 3 shows that prior familiarity had a statistically significant relationship ($p = 0.003$) on the overall perceived learning experience of students watching the video, while controlling for the background demographic and academic factors.

Table 3

Multiple Regression Output to Study the Effect of Prior Familiarity of Big Data on Students' Learning Experience

	B	SE B	β	p-value
Sex	0.119	0.134	0.263	0.037**
Age	0.018	0.018	0.074	0.314
Ethnicity: Hispanic ¹	-0.018	0.189	-0.010	0.922
Ethnicity: Asian/PI ¹	-0.102	0.196	-0.052	0.604
Ethnicity: White ¹	-0.036	0.201	-0.017	0.858
Major: Mathematics ²	0.109	0.271	0.029	0.689
Major: Health Sciences ²	0.215	0.158	0.120	0.175
Major: Biology ²	0.231	0.207	0.090	0.265
Major: Chem./Biochem. ²	0.314	0.296	0.077	0.290
Class level: Sophomore ³	-1.165	0.935	-0.505	0.214
Class level: Junior ³	-1.328	0.931	-0.739	0.155
Class level: Senior ³	-1.138	0.942	-0.564	0.229
Intended degree: Master's ⁴	0.136	0.149	0.072	0.102
Intended degree: PhD ⁴	0.150	0.157	0.276	0.023**
Prior familiarity with big data	0.311	0.226	1.223	0.003**

Notes. SE B is standard error B coefficient. ** $p < .05$, * $p < .10$.

¹Other group used as comparison for ethnicity.

²Other group used as comparison for majors.

³Freshman used as comparison for class level.

⁴Bachelor's used as comparison for intended degree.

Adjusted $R^2 = 0.336$. Model ANOVA p-value = 0.026 ($n = 238$).

The positive coefficient showed that students who were less familiar (recall that higher values of this variable indicated less familiarity) with big data before viewing the video reported having a better learning experience than those

who were more familiar. In addition, intended degree and gender had a significant effect on the overall perceived learning experience of students, which corroborates our earlier ANOVA results included in Table 1.

Interest in Pursuing the Field of Big Data Science in the Future

Data on factors influencing students’ interest in pursuing the field of big data, particularly whether prior familiarity with the topic and a positive learning experience via the video contributed to pursuing big data are shown in Table 4. Thirty-five percent of the student

participants reported that their interest in the field of big data increased after watching the video. The overall mean rating for their interest in exploring additional resources and materials on the topic was 4.42 (standard deviation = 1.645) and the mean rating for their interest in exploring career pathways in the field was 4.25 (standard deviation = 1.729). Both these variables were measured using a 6-point Likert scale as described earlier, and hence these mean values show overall positive responses from students. Table 4 shows the descriptive statistics for these two constructs by the different background demographic variables.

Table 4
Descriptive Statistics for Interest in Pursuing the Field of Big Data by Students’ Demographic and Academic Background Variables

	Interest in exploring materials		Interest in career paths	
	(Mean ± SD)	p	(Mean ± SD)	p
Overall	4.42 ± 1.65		4.25 ± 1.73	
Gender		0.556		0.274
Male	(4.50 ± 1.69)		(4.40 ± 1.65)	
Female	(4.38 ± 1.62)		(4.17 ± 1.77)	
Ethnicity		0.091*		0.004**
American Indian/Alaskan native	(3.33 ± 1.53)		(3.00 ± 1.73)	
Asian or Asian American	(4.67 ± 1.53)		(4.67 ± 1.51)	
Black or African American	(3.60 ± 1.34)		(3.80 ± 1.64)	
Hispanic	(4.57 ± 1.61)		(4.32 ± 1.79)	
Native Hawaiian/Pacific Islander	(3.60 ± 1.34)		(4.00 ± 1.41)	
White	(4.16 ± 1.85)		(3.96 ± 1.87)	
Multiethnic	(4.60 ± 1.30)		(4.40 ± 1.45)	
Other	(3.50 ± 1.83)		(2.58 ± 1.24)	
Major		0.611		0.086*
Biological Sciences	(4.64 ± 1.50)		(4.69 ± 1.56)	
Mathematics	(4.88 ± 1.90)		(4.76 ± 1.95)	
Health Sciences	(4.38 ± 1.60)		(4.13 ± 1.72)	
Psychology	(4.00 ± 1.15)		(3.75 ± 1.50)	
Biochemistry	(4.86 ± 1.51)		(5.14 ± 1.03)	
Chemistry	(4.88 ± 0.24)		(5.00 ± 0.71)	
Other	(4.43 ± 1.64)		(4.26 ± 1.72)	
Class level		0.002**		0.001**
Freshman	(5.50 ± 0.71)		(5.00 ± 0.54)	
Sophomore	(4.54 ± 1.24)		(4.34 ± 1.49)	
Junior	(4.30 ± 1.67)		(4.08 ± 1.73)	
Senior	(4.51 ± 1.89)		(4.48 ± 1.92)	
Intended Degree		0.478		0.667
Bachelor’s	(4.13 ± 1.68)		(3.89 ± 1.73)	
Master’s	(4.55 ± 1.59)		(4.38 ± 1.70)	
Ph.D., Ed.D., MD, other prof.	(4.60 ± 1.65)		(4.48 ± 1.71)	

Note. The p-values reported are from the ANOVA tests studying the difference among the different groups. *<0.10, **<0.05, ***<0.0001.

Results indicated that statistically significant differences existed among students with respect to pursuing the field of big data in the future by class level ($p = 0.001$), ethnicity ($p = 0.004$) and major ($p = 0.08$). Asians, Hispanics, and multi-ethnic groups had the highest interest in pursuing both additional materials and career pathways in this newly emerging area, while African Americans, American Indians/Alaskan Natives and other ethnic groups expressed least interest. Students with Chemistry, Biochemistry and Mathematics majors showed overall high interest in exploring this field further upon watching the video, whereas students with Psychology and Health Science majors showed relatively lower interest. Freshmen and Seniors showed overall higher interest in pursuing the field of big data than Juniors and Seniors. Men were found to have a slightly higher level of interest in pursuing the topic of big data in the future than women, although this difference was not statistically significant ($p = 0.87$).

Table 5 shows the multiple regression model to study the effect of prior familiarity with big data on pursuing the field further by exploring additional courses or career pathways. The results clearly indicated that prior familiarity had a statistically significant effect on overall learning experience ($p = 0.001$), while adjusting for other factors. The negative coefficient showed that students who were more familiar with big data before seeing the video reported higher interest in exploring the field of big data in the future than those who were more familiar. This also establishes the utility of this tutorial video as a motivating factor for students to consider a new emerging field for future studies or careers. In addition, ethnicity, class level, major, and intended degree had a significant effect on students' overall interest to explore the big data field further.

Discussion

Our findings revealed that students perceived the video as an overall very good experience and viewed it as a valuable learning tool about big data, as is evident from the mean rating of 5.74 (out of 6-point Likert scale) from over 300 students. Female students reported a

significantly better learning experience, and so did students who intended to pursue higher

Table 5

Multiple Regression of Students' Interest in Pursuing the Field of Big Data in the Future

	B	SE B	β	p-value
Sex	-0.039	0.242	-0.011	0.872
Age	0.030	0.033	0.063	0.363
Ethnicity: Hispanic ¹	0.606	0.342	0.164	0.078*
Ethnicity: Asian/PI ¹	1.092	0.355	0.288	0.002**
Ethnicity: White ¹	0.245	0.360	0.061	0.497
Major: Mathematics ²	0.912	0.491	0.123	0.065*
Major: Health Sciences ²	0.008	0.283	0.002	0.977
Major: Biology ²	0.633	0.372	0.127	0.090*
Major: Chem./Biochem. ²	1.065	0.536	0.134	0.048**
Class level: Sophomore ³	-1.326	1.691	-0.306	0.434
Class level: Junior ³	-1.570	1.687	-0.454	0.353
Class level: Senior ³	-1.232	1.707	-0.315	0.071*
Intended degree: Master ⁴	0.570	0.267	0.157	0.034**
Intended degree: PhD ⁴	0.377	0.282	0.099	0.183
Prior familiarity with big data	-0.431	0.131	-0.207	0.001**

Notes. SE B is standard error B coefficient. ** $p < .05$, * $p < .10$.

¹Other group used as comparison for ethnicity.

²Other group used as comparison for major.

³Freshman used as comparison for class level.

⁴Bachelor's used as baseline for intended degree.

Adjusted $R^2 = 0.565$. Model ANOVA p-value < 0.0001 , $n = 243$.

graduate degrees in the future. Moreover, students who had little or no familiarity with big data prior to watching the video (for instance, did not even know the word or had heard the term only without knowing what it meant and entailed) mentioned having a better learning experience than those students who had some familiarity with the topic. This strongly established the utility of this introductory video tutorial as a learning tool for first-time exposure to this important topic in today's world for a

diverse group of undergraduate students. Because females are less likely to enter STEM fields, we showed that potentially introducing data science programs to females in upper division courses, and those interested in higher education, may provide additional educational benefits for these students. Further, the video is beneficial overall to students with little familiarity, and therefore introducing the video in lower division, introductory Statistics, Computer Science and Health Science courses may promote learning for this group. Effect sizes computed using eta squared values (Richardson, 2011) showed moderate to high significance of the results (values > 0.45).

Results also showed that very few students (less than thirty percent) were familiar with the concept of big data science before watching the video beyond just knowing the term. Male students and students with Biology and Chemistry majors were more familiar with the concept of big data in terms of having some idea what it meant than others who had either not heard the term at all or just heard the term without any additional knowledge about the topic. The latter seems reasonable because those fields often give rise to large and complex datasets that students may have had exposure to in some of their previous courses. Another positive outcome of our study was that a large proportion of students who watched the video (around 65%) expressed a high level of interest in either exploring additional courses and tutorials in the area of big data or even exploring career pathways in that field, particularly those belonging to underrepresented student groups, such as females, and ethnic minority groups like Hispanics and Asians. Thus introducing the video in introductory science courses may help motivate females and underrepresented minority ethnic groups to pursue a career in big data, a concept that is unfamiliar to most prior to the video, thus narrowing the gap in the STEM workforce. Just as with the learning experience, effect sizes also showed moderate to high significance of these results (values > 0.35).

Big data is a fast-growing field, and has widespread applications today in business, healthcare, education, security, among other

fields (Mayer-Schonberger & Cukier, 2014). Although the importance of big data is already understood and concepts are being gradually integrated in college curriculum, a potentially quick and easy way to introduce a topic to a large group of college students is to use multimedia tools that make use of video, audio, animation, etc. in order to create attractive and engaging content. Multimedia has been shown to be a very effective tool for today's learners in different fields like education and science, as demonstrated by recent research studies (Aloraini, 2012, Miller et al., 2011). Babiker (2015) stated that educators must create their own multimedia applications to be effective in higher education. Our current study leveraged the potential benefits of multimedia learning in the form of a video created by faculty to introduce a diverse group of students in an undergraduate institution to the different aspects big data science, including issues and challenges that make big data different from traditional datasets that people are typically familiar with. The research goal was to understand how effective and potentially beneficial the tool was in self-perceived understanding of big data concepts and its application in health and biomedical sciences, a topic that is not studied much in the current literature, hence filling the gap.

A potential future direction of study includes looking at causal connections between student learning and multimedia based on social, affective, and cognitive factors, which has not been researched much yet. Opportunities to develop similar other multimedia tools for big data education will also be explored.

Limitations

The study was limited only to students enrolled in specific courses in Mathematics, Biology and Health Science. So, although we had an adequate sample size and relatively reasonable representation of the student population on our campus, showing the video and obtaining feedback from additional courses on campus from various disciplines might give us more accurate insights about the role of multimedia as a learning tool for a diverse undergraduate student population at our university. Further, we

did not have any knowledge of actually how many students viewed the video, we only had data on those that participated in the survey. Thus complete participation rates are unavailable and although this can potentially affect results, we do not anticipate any significant bias because the behaviors and backgrounds of the non-participants are unknown. Although one reason for non-participation maybe disinterest in the video and the topic of data science, there could potentially be several other reasons driving a student's decision to not complete the survey such as, busy schedules, technical issues (no Internet connection, say) or they simply forgot. Thus future studies should consider non-participant characteristics. Another limitation of the study is that it assessed only students' self-perceived knowledge to understand the impact of the multimedia tool in learning about big data concepts as the first step; in the future, we plan to investigate actual knowledge gains based on a pre-post type study. Finally, although many students expressed interest in pursuing the field of big data in the future, it is not known exactly how they would do so or whether they would do so at all. So it might be interesting to assess the actual future actions of such students via follow-

up surveys using the names and email addresses they had shared on the survey to be contacted in the future; nonetheless, this is beyond the scope of the present study.

Conclusion

We showed that a multimedia educational tool on the newly emerging topic of big data can be a successful tool in disseminating knowledge about big data science and its application in biomedicine among undergraduate students and creating interest for additional exploration, and that these experiences and interests varied by demographic and other factors. This can thus aid in creating a pipeline for more underrepresented students to choose careers in the field of big data science, hence narrowing the gap in the STEM workforce. We also observed that both prior familiarity with big data and overall learning experience from the video had significant effects on students' future interest in the field. However, students with less familiarity with big data prior to watching the video had better overall learning experience, whereas those with a greater level of prior familiarity with big data were more inclined to pursue the area more in future.

Acknowledgements

This study was part of the Big Data Discovery and Diversity through Research Education Advancement and Partnerships (BD3-REAP) Project funded by National Institutes of Health (NIH)-R25; Grant number is 1R25MD010397-01. The authors would like to thank Mary Aboud and Harmanpreet Chauhan for their help with the multimedia video tool, and Gaganpreet Sahota for her contribution to preliminary literature review on the topic of the effect of multimedia tools on student learning. They would also like to thank John Kucsera of Arroyo Research Services for reviewing the providing feedback on the survey questionnaire and the faculty conducting the interviews and providing their respective expertise in big data science in the video.

References

- Aloraini, S. (2012). The impact of using multimedia on students' academic achievement in the College of Education at King Saud University. *Journal of King Saud University - Languages and Translation*, 24(2), 75 - 82.
- Babiker, M.E.A. (2015). For effective use of multimedia in education, teachers must develop their own educational multimedia applications. *The Turkish Online Journal of Educational Technology*, 14(4), 62 – 68.
- Chen, H., & Xie, T. (2011). Research on the application of video cases in computer education of college students. *6th International Conference on Computer Science and Education (ICCSE)* (pp. 668 - 670). Singapore: IEEE.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297 - 334.

- Dembosky, A. (2012, December 12). *Data prescription for better healthcare*. Retrieved October 24, 2017, from Financial Times: : <http://www.ft.com/intl/cms/s/2/55cbca5a-4333-11e2-aa8f-00144feabdc0.html#axzz2W9cuwajK>
- Ed Tech. (2016). *5 benefits of multimedia learning*. Retrieved October 24, 2017, from <https://www.nuiteq.com/company/blog/5-benefits-of-multimedia-learning>
- Frost & Sullivan. (2012). *Drowning in Big Data? Reducing Information Technology complexities and costs for healthcare organizations*. Retrieved October 24, 2017, from <http://www.emc.com/collateral/analyst-reports/frost-sullivan-reducing-information-technology-complexities-ar.pdf>
- Holmes, D.E. (2018). *Big Data: A Very Short Introduction*, 1st edition, Oxford University Press.
- Huang, T., Lan, L., Fang, X., An, P., Min, J., & Wang, F. (2015). Promises and challenges of Big Data computing in health sciences. *Big Data Research*, 2(1), 2 - 11.
- IBM SPSS Website: <http://www-01.ibm.com/support/docview.wss?uid=swg24041224>. Accessed December 2017.
- Ilan, G. O., & Oruc, S. (2016). Effect of the use of multimedia on students' performance: A case study of social studies class. *Academic Journals*, 18(8), 877 - 882.
- Liaskos, J., Diomidus, M. (2002). Multimedia technologies in education, *Studies in Health Technology and Informatics*, 65, 359 – 372.
- Marr, B. (2015, September 30). *Big Data: 20 Mind-boggling facts everyone must read*. Retrieved October 24, 2017, from Forbes: <https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#210f754c17b1>
- Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, 81(2), 240–246.
- Mayer-Schonberger, V. & Cukier, K. (2014). Big data: A revolution that will transform how we live, work, and think. *John Murray publishers, UK*.
- Miller, L. M., Chang, C. -I., Wang, S., Beier, M. & Kilsch, M. (2011). Learning and motivational impacts of a multimedia science game. *Computers and Education*, 57, 1425 – 1433.
- Office of Technology Assessment, U.S. Congress. (1988). *Power on! New tools for teaching and learning: A new course for education*. Washington DC: U.S. Government Printing Office.
- Office of Institutional Research and Analytical Studies, California State University, Fullerton. Student Demographics – Headcount by Gender and Ethnicity. Website: <http://www.fullerton.edu/analyticalstudies/facts/demographics.php>. Accessed on June 1, 2018.
- Richardson, T.E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, 6(2), 135 – 147.
- Tang, S. (2014). The Impact of Multimedia on Social Learning. *Journal of Communication and Education*, 1(1), 39 - 45.
- Yang, D. (2010). Discusses in shallowly the classroom instruction the video frequency case application. *The Guide of Science and Education*, 15.

Author Information

Sinjini Mitra, Ph.D. (corresponding author)
Department of Information Systems & Decision Sciences
California State University, Fullerton
800 N. State College Blvd., SGMH – 4113, Fullerton, CA
92831.
Phone: (657) 278-2234
Email: smitra@fullerton.edu