

Big data and student engagement among vulnerable youth: A review

Ryan J Watson^{1,3} and John L Christensen^{2,3}



Vulnerable (e.g., LGBTQ, homeless, disabled, racial/ethnic minority, and/or poor) youth disproportionately report challenges at school compared to their majority counterparts, but we are not always sure of the best ways to support these students. How might big data help to ameliorate experiences for vulnerable students who are not part of the majority (e.g., White, middle class, straight)? We review current ways that using big data can promote student engagement specific to school experiences where vulnerable youth share a disproportional amount of burden. We review extant uses of big data to track, involve, and monitor student progress and attendance. Additionally, we review the potential privacy implications and threats to students' civil liberties.

Addresses

¹ Department of Human Development and Family Studies, University of Connecticut, 348 Mansfield Road, U-1058, Storrs, CT 06269, United States

² Department of Communication, University of Connecticut, Storrs, CT, United States

Corresponding author: Watson, Ryan J (ryanwatson@uconn.edu)

³ The authors contributed equally to the work of this manuscript.

Current Opinion in Behavioral Sciences 2017, 18:23–27

This review comes from a themed issue on **Big data in the behavioural sciences**

Edited by **Michal Kosinski** and **Tara Behrend**

<http://dx.doi.org/10.1016/j.cobeha.2017.07.004>

2352-1546/© 2017 Elsevier Ltd. All rights reserved.

Introduction

While research on school experiences for students has burgeoned over the past several decades [1–4], we know less about the experiences of vulnerable young people, especially within the school system, with few exceptions [5,6^{**}]. Here, we operationalize vulnerable youth as those with stigmatized and disadvantaged identities or positions in society; for example, LGBTQ, disabled, homeless, racial/ethnic minority, and/or poor young people. Research has found that many youth part of these vulnerable groups feel unsafe at school, which, for LGBTQ and racial minority youth as an example, is largely attributed to high rates of bias-based victimization [7,8]. For many of these youth, school environments are not becoming safer over time [6^{**}];

there is foundational evidence that vulnerable groups of youth report more hostile school climates, greater absenteeism, compromised academic achievement, and lower academic aspirations compared to White, heterosexual, and able-bodied young people [9]. To address these clear disparities, scholars of the future will need to leverage novel techniques, such as the use of big data.

Big data in the social sciences has been operationalized as an interplay between technology, mythology, and analysis; that is, analyzing, aggregating, and cross-referencing large datasets [10]. While big data has enhanced our ability to reach the masses in new and exciting ways, vulnerable youth (especially young people in schools) in particular have been difficult to study for some time; these youths may have invisible identities (e.g., LGBTQ youth), be hard to reach (e.g. homeless youth), or not be available for survey participation (e.g., disabled youth). In this review of big data in school settings, we focus on vulnerable youths' experiences at school, where they are oftentimes burdened with a disproportional amount of bullying and risk. It is important to note that providing recommendations for the practical application of big data in school settings is not the intention of this article. Instead, our goal is to provide a review of recent literature so that readers will be informed about practices being discussed by big data scholars.

Role of big data for vulnerable youth in schools: Student engagement

Information available to school administrators has traditionally been limited to variables such as conduct history and academic achievement, but novel sources of big data offer expanded possibilities relevant to the bullying and safety concerns faced by vulnerable students [11]. Modern advances in the collection and analysis of big data afford opportunities to draw insights from information that has previously been unavailable or difficult to access [10]; such information is currently being leveraged to improve the academic environment [12,13] in ways that benefit a wide range of stakeholders, including students who are vulnerable to bullying [6^{**},7], administrators and teachers invested in creating a safer school environment for students [14–17], and caregivers invested in the success, health, and happiness of their children [18]. In particular, big data might play a role in influencing *student engagement* [19].

Student engagement is typically conceptualized as a multidimensional construct; a predominant model [20] operationalizes student engagement as three distinct

dimensions: cognitive, behavioral, and emotional engagement. Cognitive engagement involves self-regulated learning and the use of deep learning strategies; behavioral involves participation, effort, and conduct issues such as absenteeism; and emotional involves feelings of belongingness and the student's interpersonal relationships with teachers, classmates, and other members of the school community. Here, we focus on the latter two dimensions: Behavioral and emotional engagement are fruitful to explore in relation to vulnerable youth, given that vulnerable youth often experience higher rates of absenteeism [21], mental health concerns [22], and school victimization [23] compared to majority youth.

Behavioral engagement

Although numerous determinants of behavioral engagement have been identified, one aspect is particularly relevant to vulnerable youth: *absenteeism* (both excused and unexcused absences).

Absenteeism

Typical strategies employed to address absenteeism are parental and administrator involvement, yet vulnerable students oftentimes report poor parental, teacher, and administrator relationships [2]. Vulnerable youth report high rates of absenteeism at school [21], which research has linked to poor academic performance [24,25] and higher rates of school dropout [26], among other issues [27,28]. Vulnerable students have higher rates of absenteeism in part because they skip school to avoid getting bullied and harassed by their peers [21]. Clearly, these students are hesitant to engage with a culture that remains hostile, and removing oneself from a negative school environment serves as a coping strategy to keep one safe. Accordingly, absenteeism might be conceptualized as an indicator of bullying behaviors. In order to improve the culture within a school, administrators could use absenteeism data to better understand where bullying is occurring, when, and by whom.

School administrators are increasingly relying on digital *dashboards* in an effort to track absenteeism [29]. Dashboards are electronic interfaces that aggregate data stored in student information systems. Real-time summaries are produced, which can be used as decision aids by school administrators [30]. To facilitate ease of use, these dashboards typically display simplified statistical visualizations of key metrics, often in the form of historical trends. Data regarding excused and unexcused absences can be analyzed at multiple levels — school wide trends versus individual student trends, for instance. Moving beyond post hoc truancy monitoring, administrators can take preventative action by employing predictive modeling techniques that mine data and forecast outcomes [31]. These dashboard systems are able to monitor and predict absenteeism in a granular way, with predictability enhanced by the quantity and

quality of input variables available when computing an estimated risk-factor score.

Given that many student-oriented events take place outside of the immediate school environment [32], additional big data technologies are relevant. For example, location-based data sources have been used to estimate and encourage participation. This may include self-reported location 'check-ins' on social media [33] or location data captured by smartphones equipped with GPS technology [34]. Even data obtained from radio-frequency identification tags (RFID) [35] and vehicle license plate readers [36] have been deemed useful sources of information by some. Another source of data comes from student identification cards [19]. Commonly used at colleges and universities, many school administrators require these cards to be swiped whenever a student enters a residential building, dining hall, library, or sporting venue. Although potentially useful to administrators, these applications of big data raise privacy concerns, which are discussed below.

Emotional engagement

In the context of vulnerable students' experiences, two aspects of emotional engagement are especially relevant. The first involves students' emotional reactions to educational activities, their teachers, and their peers. The second involves the feelings of belongingness and relatedness that vulnerable students may or may not be experiencing within the school setting.

School environments can elicit a host of emotional reactions; some (e.g., hope) have been linked to positive learning outcomes such as self-regulated learning and achievement whereas others (e.g., boredom) have been found to derail success [37]. Given the mental health disparities experienced by vulnerable youth [38], administrators have attempted to assess their school's emotional climate and intervene when necessary [39]. For example, some are using protective software that allows them to monitor social media posts submitted by students [40]. Sophisticated algorithms identify linguistic patterns in the data that are predictive of emotional distress, mood states, and mental health problems [41], such as depression. When warning signs are evident, administrators may choose to intervene with the goal of eliminating existing safety concerns and preventing new ones before they arise.

Emotional engagement extends to other components of one's school experience such as involvement in related activities. For instance, participation in school sporting events may serve to increase one's identification with peers and perceptions of being an important and valuable member of the school community [42]. Similarly, emotional engagement is facilitated by the positive emotional experiences that stem from being an active member of

groups focused on student governance, professional development, community service, and other special interests [43]. Groups particularly relevant to the population at hand include gay/straight alliances and political advocacy committees. Because participation in these groups may serve to increase feelings of belongingness and relatedness, school administrators may choose to implement some of the same technologies used to prevent absenteeism.

In addition to these technologies, other emerging sources of big data might be leveraged in an effort to predict vulnerable students' feelings and sense of belonging. Two examples involve insight gleaned from social networking sites such as Facebook and Twitter. The first is social network analysis [44,45^{*}], which is the process of visualizing the nodes and ties that form complex social structures (i.e. the student body) [46]. The number, mutuality, and strength of one's interpersonal connections could possibly be used to predict emotional engagement and intervene at the school level when students are experiencing low levels of belonging. A second emerging technology focuses on the analysis of user-generated text. Sentiment analysis [47^{**}] is being utilized in some schools to detect and quantify the emotional content of phrases — whether they are written in documents, published on blogs, or posted to social networking sites. Basic algorithms assist school administrators in the detection of positive, negative, or neutrally valenced words and phrases. Advanced applications have been developed to gain more nuanced insight relevant to the detection of aggression, hostility, and fear related to bullying [48]. The data obtained has the potential to provide an understanding of students' feelings of belongingness and relatedness with peers and the school community at large.

Privacy concerns: Big data and the protection of civil liberties

As big data technologies advance, opportunities for bullying prevention, detection, and intervention will surely grow in tandem. But, along with these benefits come significant privacy implications that cannot be ignored. The potential for violating civil liberties when implementing big data strategies should be of utmost concern [49–52]. And this is particularly concerning when the population includes youth (students) who are members of vulnerable populations subject to unwanted scrutiny (e.g., undocumented immigrants, LGBTQ). For example, it is a very real possibility that vulnerable students facing social and emotional stressors may further withdraw from the school environment when they are informed that school administrators are monitoring their behavioral and emotional engagement via big data. Specific to students identifying as LGBTQ, what might happen if in the process of preventing bullying, those who are sexual minorities are 'outed'? And regarding students who are dealing with substance abuse issues, could their data be obtained by police and used to prosecute them

[53]? In an effort to maintain anonymity, students may purposefully limit what they say and do, or refrain from reaching out for help and support, which may detract from their well-being and social development [54].

A recent study that explored privacy expectations and experiences among marginalized Internet users in a classroom setting found students had low expectations of privacy and high expectations of surveillance [55^{**}]; these students reported a lack of access to adequate solutions. In line with these concerns, the United States Department of Education has established a set of requirements and best practices for protecting student privacy in the digital age, such as being transparent with students, considering that parental consent may be appropriate, and maintaining awareness of laws like FERPA that are designed to protect student information. Despite these efforts, big data scholars have recently argued that considerable confusion remains and new student privacy laws are needed [56,57].

Conclusion and additional drawbacks of big data

Despite the clear strengths in using big data to better engage vulnerable students at school, there are some additional drawbacks of a practical nature that must be acknowledged. For example, analysis of big data is a complex process that requires an investment in financial and human resources [13]. Though the potential benefits for youth may be welcomed by relevant stakeholders, the implementation of big data analysis may be slow given these restrictions. Furthermore, any conclusions drawn from big data must be interpreted with caution due to methodological challenges such as a reliance on biased sampling frames and limited or incomplete data [58].

In sum, even some of the today's most challenging student engagement problems may be easily addressable by leveraging big data, yet the potential drawbacks are meaningful. The ability of school administrators to improve the lives of vulnerable students via big data seems promising — if implemented with careful consideration of this special population's civil liberties.

Conflict of interest statement

Nothing declared.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Connell JP, Spencer MB, Aber JL: **Educational risk and resilience in African-American youth: context, self, action, and outcomes in school.** *Child Dev* 1994, **65**:493-506.
2. Gilligan R: **Adversity, resilience and young people: the protective value of positive school and spare time experiences.** *Child Soc* 2000, **14**:37-47.

3. Bryan J: **Fostering educational resilience and achievement in urban schools through school-family-community partnerships.** *Prof Sch Counsel* 2005, **8**:219-227.
4. Cappella E, Weinstein RS: **Turning around reading achievement: predictors of high school students' academic resilience.** *J Educ Psychol* 2001, **93**:758-771.
5. Juvonen J, Graham S (Eds): *Peer Harassment in School: The Plight of the Vulnerable and Victimized.* Guilford Press; 2001.
6. Goodenow C, Watson RJ, Adjei J, Homma Y, Saewyc E: **Sexual orientation trends and disparities in school bullying and violence-related experiences, 1999–2013.** *Psychol Sex Orient Gender Divers* 2016, **3**:386-396.
- Examines trends and disparities in bullying in schools for heterosexual and sexual minority adolescents. The authors found that rates of bullying are increasing for some subgroups of sexual minorities, and the disparities are getting worse for sexual minority girls.
7. Toomey RB, Ryan C, Diaz RM, Card NA, Russell ST: **Gender-nonconforming lesbian, gay, bisexual, and transgender youth: school victimization and young adult psychosocial adjustment.** *Dev Psychol* 2010, **46**:1580-1589.
8. Lee VE, Burkam DT: *Inequality at the Starting Gate: Social Background Differences in Achievement as Children Begin School.* Washington, DC: Economic Policy Institute; 2002.
9. Pearson J, Muller C, Wilkinson L: **Adolescent same-sex attraction and academic outcomes: the role of school attachment and engagement.** *Soc Probl* 2007, **54**:523-542.
10. Boyd D, Crawford K: **Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon.** *Inf Commun Soc* 2012, **15**:662-679.
11. Sin K, Muthu LL: **Application of big data in education data mining and learning analytics – A literature review.** *ICTACT J Soft Comput* 2015, **5**:1-35.
12. Selwyn N, Henderson M, Chao SH: **Exploring the role of digital data in contemporary schools and schooling – '200,000 lines in an Excel spreadsheet'.** *Br Educ Res J* 2015, **41**:767-781.
13. Daniel B: **Big data and analytics in higher education: opportunities and challenges.** *Br J Educ Technol* 2015, **46**:904-920.
14. Cabellon ET, Junco R: **The digital age of student affairs.** *New Direct Stud Serv* 2015, **151**:49-61.
15. Williamson B: **Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments.** *J Educ Policy* 2016, **31**:123-141.
16. Timms MJ: *Big Data in Education: A Guide for Educators.* Centre for Strategic Education; 2015.
17. Selwyn N: **'There's so much data': exploring the realities of data-based school governance.** *Eur Educ Res J* 2016, **15**:54-68.
18. Abreu A, Rocha Á, Cota MP: **Perceptions of teachers and guardians on the electronic record in the school-family communication.** In *Conference on e-Business, e-Services and e-Society.* Edited by Janssen M. Springer International Publishing; 2015:48-62.
19. Gross L, Meriwether JL: **Student engagement through digital data.** *New Direct Stud Serv* 2016, **155**:75-89.
20. Fredricks JA, Filsecker M, Lawson MA: **Student engagement, context, and adjustment: addressing definitional, measurement, and methodological issues.** *Learn Instruct* 2016, **43**:1-4.
21. Burton CM, Marshal MP, Chisolm DJ: **School absenteeism and mental health among sexual minority youth and heterosexual youth.** *J Sch Psychol* 2014, **52**:37-47.
22. Marshal MP, Dietz LJ, Friedman MS, Stall R, Smith HA, McGinley J, Thoma BC, Murray PJ, D'Augelli A, Brent DA: **Suicidality and depression disparities between sexual minority and heterosexual youth: a meta-analytic review.** *J Adolesc Health* 2011, **49**:115-123.
23. Toomey RB, Russell ST: **The role of sexual orientation in school-based victimization: a meta-analysis.** *Youth Soc* 2016, **48**:176-201.
24. Lamdin DJ: **Evidence of student attendance as an independent variable in education production functions.** *J Educ Res* 1996, **89**:155-162.
25. Dembo R, Wareham J, Schmeidler J, Briones-Robinson R, Winters KC: **Differential effects of mental health problems among truant youths.** *J Behav Health Serv Res* 2016, **43**:402-427.
26. Lee JS: **The relationship between student engagement and academic performance: is it a myth or reality?** *J Educ Res* 2014, **107**:177-185.
27. Robins LN, Ratcliff KS: **The long-term outcome of truancy.** In *Out of School: Modern Perspectives in Truancy and School Refusal.* Edited by Hersov LA. Wiley & Sons; 1980:65-83.
28. Russell ST, Ryan C, Toomey RB, Diaz RM, Sanchez J: **Lesbian, gay, bisexual, and transgender adolescent school victimization: implications for young adult health and adjustment.** *J Sch Health* 2011, **81**:223-230.
29. West DM: **Big data for education: data mining, data analytics, and web dashboards.** *Gover Stud Brook* 2012, **4**:1-10.
30. Clune B, Knowles J: *Delivery of state-provided predictive analytics to schools: Wisconsin's DEWS and the proposed EWIMS dashboard (WCER Working Paper No. 2016-3).* University of Wisconsin-Madison, Wisconsin Center for Education Research; 2016.
31. Rajni J, Malaya DB: **Predictive analytics in a higher education context.** *IT Prof* 2015, **17**:24-33.
32. Li Y, Bebiroglu N, Phelps E, Lerner RM, Lerner JV: **Out-of-school time activity participation, school engagement and positive youth development: findings from the 4-H Study of Positive Youth Development.** *J Youth Dev* 2016, **3**:1-22.
33. Sims KM, Weber EM, Bhaduri BL, Thakur GS, Resseguie DR: **Application of social media data to high-resolution mapping of a special event population.** In *Advances in Geocomputation.* Edited by Griffith D, Chun Y, Dean D. Springer; 2017:67-74.
34. Chow PI, Fua K, Huang Y, Bonelli W, Xiong H, Barnes LE, Teachman BA: **Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students.** *J Med Internet Res* 2017, **19**:e62.
- Tested the hypothesis that more time spent at home would be associated with higher levels of negative affect and lower levels of positive affect among college students. GPS location data was collected by an app installed on the students personal mobile phones. The findings provide evidence of the feasibility and utility of this innovative modelling approach.
35. Anjum R, Kamble V: **RFID student navigation system.** *Int J Sci Res Sci Technol* 2016, **2**:463-467.
- Describes the design and implementation of an automatic student attendance system based on radio frequency identification (RFID) technology. The authors conclude that the system is effective and more convenient than manual attendance tracking systems that educational institutions have traditionally employed.
36. Zmud J, Wagner J, Moran M: **License plate reader technology.** *Transp Uses Priv Risks Natl Coop Highw Res Prog* 2016:1-90.
37. Pekrun R, Goetz T, Titz W, Perry RP: **Academic emotions in students' self-regulated learning and achievement: a program of qualitative and quantitative research.** *Educ Psychol* 2002, **37**:91-105.
38. *Institute of Medicine: The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a Foundation for Better Understanding.* Washington, DC: Institute of Medicine; 2011.
39. Sellar SL: **A feel for numbers: affect, data and education policy.** *Crit Stud Educ* 2015, **56**:131-146.
40. Shade LR, Singh R: **"Honestly, We're Not Spying on Kids": school surveillance of young people's social media.** *Soc Media Soc* 2016, **2**:1-12.
41. Locke C: *How Analyzing Data Can Prevent Cyberbullying and Suicide.* 2014. Retrieved from: <https://www.edsurge.com/news/>

- 2014-10-29-how-analyzing-data-can-prevent-cyberbullying-and-suicide.
42. Van Boekel M, Bulut O, Stanke L, Zamora JR, Jang Y, Kang Y, Nickodem K: **Effects of participation in school sports on academic and social functioning.** *J Appl Dev Psychol* 2016, **46**:31-40.
 43. Martinez A, Coker C, McMahon SD, Cohen J, Thapa A: **Involvement in extracurricular activities: identifying differences in perceptions of school climate.** *Educ Dev Psychol* 2016, **33**:70-84.
 44. Lee J, Bonk CJ: **Social network analysis of peer relationships and online interactions in a blended class using blogs.** *Internet High Educ* 2016, **28**:35-44.
 45. DeLay D, Hanish LD, Zhang L, Martin CL: **Assessing the impact of homophobic name calling on early adolescent mental health: a longitudinal social network analysis of competing peer influence effects.** *J Youth Adolesc* 2016, **46**:955-969.
 Utilizes longitudinal social network analysis in an effort to improve understanding of the relationship between homophobic name calling and middle school students' mental health. An assessment of peer relationship networks provided evidence of a significant impact of homophobic name calling on indicators of mental health. The article discusses the role of social network analysis in the implementation of effective intervention and prevention strategies.
 46. Dawson S, Bakharia A, Lockyer L, Heathcote E: **'Seeing' networks: visualising and evaluating student learning networks-report.** *Education* 2011, **3**:1-6.
 47. Esparza GG, Díaz AP, Canul-Reich J, De Luna CA, Ponce J: **Proposal of a sentiment analysis model in tweets for improvement of the teaching-learning process in the classroom using a corpus of subjectivity.** *Int J Combinat Opt Probl Inf* 2016, **7**:22-34.
 Proposes a sentiment analysis model with the goal of assessing teacher performance based on Twitter posts submitted by college students. Naïve Bayes (NB) is described as a technique that enables the classification of positive, neutral, and negative content in an educational assessment context.
 48. Rafiq RI, Hosseinmardi H, Mattson SA, Han R, Lv Q, Mishra S: **Analysis and detection of labeled cyberbullying instances in Vine, a video-based social network.** *Soc Netw Anal Min* 2016, **6**:88.
 49. Selwyn N: **Data entry: towards the critical study of digital data and education.** *Learn Media Technol* 2015, **40**:64-82.
 50. Trainor S: **Student data privacy is cloudy today, clearer tomorrow.** *Phi Delta Kappan* 2015, **96**:13-18.
 51. Scholes V: **The ethics of using learning analytics to categorize students on risk.** *Educ Technol Res Dev* 2016, **64**:939-955.
 52. Slade S, Prinsloo P: **Learning analytics: ethical issues and dilemmas.** *Am Behav Sci* 2013, **57**:1510-1529.
 53. Crawford K, Schultz J: **Big data and due process: toward a framework to redress predictive privacy harms.** *BCL Rev* 2014, **55**:93.
 54. O'Neil C: *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown Publishing Group; 2016.
 55. Gangadharan SP: **The downside of digital inclusion: expectations and experiences of privacy and surveillance among marginal Internet users.** *New Media Soc* 2017, **19**:597-615.
 This mixed-methods study focused on marginalized students and demonstrated that they had high expectations of Internet surveillance, were exposed to potential harms such, and had little access to adequate solutions.
 56. Weber AS: **The big student big data grab.** *Int J Inf Educ Technol* 2016, **6**:65.
 57. Weber A: **Suggested legal framework for student data privacy in the age of big data and smart devices.** *Front Artif Intell Appl* 2014, **262**:669-678.
 58. Hargittai E: **Is bigger always better? Potential biases of big data derived from social network sites.** *Ann Am Acad Polit Soc Sci* 2015, **659**:63-76.