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 disproportionally 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.

Introduction

While research on school experiences for students has bourgeoned 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