IDHE REPORT

Creating and Maintaining the National Study of Learning, Voting, and Engagement Database

BY NANCY THOMAS AND JODI BENENSON

BASED AT TUFTS UNIVERSITY'S JONATHAN M. Tisch College of Civic Life, the Institute for Democracy and Higher Education (IDHE) works to advance learning that strengthens democracy and advances social and political equity. IDHE's signature initiative, the National Study of Learning, Voting, and Engagement (NSLVE), was established in 2013 as both a service to higher education – providing participating colleges and universities with their students' registration and voting rates – and a significant database for research. The NSLVE database is created by matching publicly available local and state voting records with enrollment lists from colleges and universities nationally. Institutions must opt in, and at the time of this writing, the study includes more than 1,000 Title IV, degree-granting colleges and universities. These institutions represent various types (e.g., community colleges, liberal arts colleges), missions (e.g., religiously affiliated institutions), student populations (e.g., full-time, part-time), and geographic locations. All 50 states are represented in the database. The database consists of enrollment records for approximately 10 million students from the relevant fall semester and voting records from the 2012 and 2016 presidential elections and the 2014 midterm election. The NSLVE data consists of de-identified student records and contains no names or other information that would allow researchers to identify an individual student.

In this report, we describe the systematic process of creating and maintaining the NSLVE database. The process includes: 1) recruiting college and university campuses to obtain permission to use student data, 2) partnering with the National Student Clearinghouse ("the Clearinghouse") to obtain student enrollment

records, 3) purchasing publicly available voting records from an organization called Catalist, and 4) working with the Clearinghouse to merge and de-identify student enrollment and voting records, and 5) calculating institutional voting rates.

1. Recruiting Colleges and Universities

Participation in NSLVE is not automatic, and colleges and universities must opt into the study by signing an authorization form allowing their enrollment records to be used for NSLVE specifically. The students in the NSLVE database are those who were on enrollment lists of participating institutions on a date closest to but before the November elections in 2012, 2014, and 2016.

To participate in the study, institutions must be degree-granting, not-for-profit public and private institutions in the U.S., and they must provide enrollment records to the Clearinghouse. To recruit campuses, we formed partnerships with existing associations and consortia in higher education, such as the American Association of State Colleges and Universities (AASCU), the Association for Institutional Research (AIR), Campus Compact, and the Higher Education Data Sharing Consortium (HEDS). We sought a range of partnerships to reach institutions of all Carnegie Classifications (state colleges, community colleges) and individuals on campus who might be interested in the data (institutional researchers and civic engagement offices). We asked these partners to disseminate information about the study to their members and networks. For many partners, we hosted informational webinars and teleconferences to answer questions about the study. We also created a "Frequently Asked Questions" about NSLVE, http://activecitizen.tufts.edu/nslve-faq/, as well as a separate FAQ on student privacy protections, http://activecitizen.tufts.edu/wp-content/uploads/NSLVE-FERPA-FAQ.pdf.

Participation in the study is free, and each participating institution receives a tailored report containing that institution's student voter registration and voting rates. One recruiting goal is for the study to obtain a representative sample of colleges and universities in the U.S.

2. Partnering with the National Student Clearinghouse

Founded in 1993, the Clearinghouse was established to streamline the process of allowing lenders to confirm that students with loans were still in school and therefore eligible for deferment of repayment. Over time, the Clearinghouse's role expanded to include degree verification, research, and other services. Currently a vast majority of U.S. colleges and universities participate in the Clearinghouse, and students at these institutions represent over 98% of students enrolled at private and public U.S. institutions. Students not in the NSLVE database are usually those who have exercised their right pursuant to the Family Education Rights to Privacy Act (FERPA) to deny the institution the right to use their information for any purpose, including for institutional research.

To participate in the Clearinghouse, institutions must provide specific information on each student, including name, date of birth, last known permanent address, and

enrollment status. Institutions have the option of providing information such as major field of study (e.g., physics, business), class level (e.g., sophomore, graduate student), race/ethnicity, gender, and whether the student is seeking a degree. In 2012, the Clearinghouse started using official Classification of Instructional Programs (CIP) codes to capture a student's field of study (NCES, 2010). Some types of institutions, such as two-year private and for-profit institutions, are less likely than others to provide data to the Clearinghouse. Dynarski, Hemelt, and Hyman (2015) find that enrollment coverage is highest among public institutions and lowest among for-profit colleges. In addition, not all colleges and universities report student characteristics such as race/ethnicity and gender. Dynarski et al. (2015) found that enrollment coverage is lower for minorities but similar for males and females. Coverage also varies in the NSLVE dataset by field of study, enrollment status (full-time or part-time), and degree seeking status.

Use of the Clearinghouse data by policy makers, educators, and academic researchers is relatively new, and because the Clearinghouse data is at the student level, it offers a significant contribution to education research. Researchers have previously used Clearinghouse data to explore the effects of particular programs or policies on postsecondary attendance, persistence, and attainment (Dynarski, Hemelt, & Hyman, 2013; Dynarski, Hyman, & Schanzenbach, 2013; Hyman, 2013). For many years, the U.S. Department of Education has been seeking legislative permission to augment the existing Integrated Postsecondary Education Data System (IPEDS), which collects institution-level data, with student-level data to enable research using individual student records.

Table 1 provides information on variables from the Clearinghouse that are included in the NSLVE dataset. About one-third of NSLVE institutions provide race/ethnicity information for their students. Race/ethnicity is particularly important to NSLVE because, in completing that data field, institutions are given the opportunity to identify nonresident aliens, generally international students who are non-citizens. Also, the Clearinghouse recently began requiring colleges and universities to identify each student as either a graduate or undergraduate student. Previously, to distinguish between graduate and undergraduate students, we relied on "class level" information, which was not reported for all students. The NSLVE database contains no student names or data that would allow researchers to identify any student.

Table 1. NSLVE Variables from National Student Clearinghouse

Variable Name	Description
OPE ID	Office of Postsecondary Education ID.
Unique ID	Generated ID number identifying each record in the file (specific to student/campus).
NSC link	Generated ID number, identifying each student in the file (can be used to match students across campuses).
Birth year	Year of the student's birth.

Variable Name	Description
Age at election	Age at election is computed from the campus-provided birthdate.
Campus city	City where the campus is located.
Campus state	State where the campus is located.
Home zip code	Earliest known zip code for the student.
Class	Class level for the student at the time of the election. Options include: Associate's, Bachelor's, Freshman, Sophomore, Junior, Senior, Undergraduate Certificate, Unspecified undergraduate, Master's, Doctoral, Post-doctorate, First professional, Unspecified graduate/sessional, Post baccalaureate certificate.
Program level+	Program level for the student at the time of the election. If more than one was provided, the highest level was selected. Options include: Undergraduate Certificate, Associate's Degree, Bachelor's Degree, Post Baccalaureate Certificate, Master's Degree, Doctoral Degree, First Professional Degree, Graduate/Professional Certificate, Non Credential Program (Preparatory Coursework/Teacher Certification).
Major	Major for the student at the time of the election.
CIP code	Classification of Instructional Program (CIP) code at the time of the Election. This may not be associated with the major reported in analysis file. If no CIP was available from the institution, NSC imputed the code from the free-text major field.
CIP description	NCES description of the CIP code.
CIP family	NCES program area to which the CIP code belongs.
CIP imputed flag	Y = The CIP code was imputed from the free-text major. Null = The CIP code is as provided by the participating campus.
Race*	Race for the student at the time of the election. Options include Nonresident alien, Asian, Black, American Indian/Alaskan Native, Asian/Pacific Islander, Hispanic, Native Hawaiian or Other Pacific Islander, White, Two or More Race/Ethnicity Categories, Race/Ethnicity Unknown. Only available for campuses that signed non-directory form.
Gender*	Gender for the student at the time of the Election. Options include Male, Female. Only available for campuses that signed non-directory form.
Enrollment status*	Enrollment status for the student at the time of the Election. Options include Full-time, Part-time (Quarter-time, half-time, or less-than-halftime status). Only available for campuses that signed non-directory form.

Variable Name	Description
Degree seeking*	Degree-seeking indicator for the student at the time of the Election. Options include: Student is seeking a degree, Student is not seeking a degree. Only available for campuses that signed non-directory form.
Age flag	Eligibility based on age. Options include: Yes = campus-provided birth date is before or equal to November 6, 1994 or November 4, 1996, No= campus-provided birth date is after November 6, 1994 or November 4, 1996, U= birth date unavailable.
SSN flag	Eligibility based on Social Security Number (SSN). Options include Yes = campus provided a valid SSN for this student, N= campus did not provide a valid SSN for this student.

* Provided for only institutions that sign an authorization form allowing permission for NSC to share this "non-directory" data element

+Only available in 2014 and 2016

3. Purchasing Voting Records from Catalist

The NSLVE database uses publicly available state and local voting records collected by Catalist, an organization that collects, cleans, and updates voter files of more than 180 million registered voters in all 50 states and the District of Columbia. Whether a person registered to vote and voted (not for whom they voted) are matters of public record, but because voting records are inconsistently maintained by states and municipalities, they can be challenging to track down. Catalist sells subscriptions to organizations interested in using the database to conduct research and is widely respected and used by academic researchers (see Ashok, Feder, McGrath, & Hersh, 2014; Ansolabehere & Hersh, 2012, 2013). While Catalist does its own modeling for clients, we use only the publicly available voting records the organization collects.

Pursuant to a contract with Tufts University, Catalist creates a snapshot of the complete set of voting records for each federal election. This snapshot is preserved for use as more colleges and universities join the study. Snapshots are created once all of the voting records have been compiled after an election (to date, April 2013, April 2015, and June 2017). Catalist has created three snapshots, one for each of the 2012, 2014, and 2016 elections. These snapshots are solely for the purpose of this study and are not accessible to others. Table 2 lists NSLVE variables that come from Catalist.

Table 2. NSLVE Variables from Catalist

Variable Name	Description
Match confidence	Match confidence score indicating the level of confidence that student in Clearinghouse database was correctly matched to voting records in Catalist database.
Election day age	Age at election is computed from the campus-provided birthdate.
Gender	Gender represented as male, female, or unknown. From voter file.
Race	Possible values: Black, Caucasian, Asian, Native American, Middle Eastern, Hispanic, Jewish, Other, Unknown. May come from voter file or a Catalist-generated algorithm predicting race based on commercial and other data sources.
Race confidence	Possible values: Possibly, Likely, Highly Likely, Uncoded. Level of confidence that Catalist's algorithm predicted the correct race/ethnicity.
Ethnicity	Ethnic subcategory breakdown. This is always supplied by the commercial vendor and is provided as a self-explanatory decoded string.
Registration date	The date the student registered to vote on the current file. Always comes from the voter file, but may have 00 for day and/or month.
Voter status	Options: 'active' = active voter registration, 'inactive' = only set for people on the current voter file with an inactive value set in that file, 'dropped' = individual has vote history and appeared on past files but does not appear in the most recent voter file, 'multiple appearances' = person with a mailing address in that state, who was sourced from another state file, and who is not found as re-registered for that state, 'unregistered' = voter is not on the current or past voter files but is known to reside in the state, 'unmatched member' = record uploaded by the client that did not match the existing Catalist database.
Registration address/city	Field is a subset of the CASS (Coding Accuracy Support System) address fields split into component parts.
Registration address/state	Valid postal abbreviation for state.

Variable Name	Description
Registration address/zip code	5 or 9 digit zip when known.
Mailing address/city	Field is a subset of the CASS parsed address fields split into component parts.
Mailing address/state	Valid postal abbreviation for state.
Mailing address/zip	5 or 9 digit zip when known.
Mailing address/state	Valid postal abbreviation for state.
County FIPS	Valid 3-digit county FIPS (Federal Information Processing Standard) code.
Method of voting - general election	Possible values: absentee, early vote, mail, polling, unknown
Precinct code	Definition, usage, and fill rate may vary by state.
Ward	Appended snapshot. Definition, usage, and fill rate may vary by state.
Congressional district	Current congressional district for voter, as supplied by voter file. When not present on the voter file, this may be imputed from the registration address.
House district	Current state house district for voter, as supplied by voter file. When not present on the voter file, this may be imputed from the registration address.
Senate district	Current state senate district for voter, as supplied by voter file. When not present on the voter file, this may be imputed from the registration address.
City council	Definition, usage, and fill rate may vary by state.

4. Merging and De–identifying Student Enrollment and Voting Records

The Clearinghouse performs the task of running the algorithm created by Catalist to match enrollment and voting records. As noted by Berent, Krosnick, and Lupia (2011), because private firms such as Catalist make earnings from proprietary models, there is potentially less transparency in the validation process when working with an outside company. For this reason, we consulted with Catalist on several occasions and turned to Ansolabehere and Hersh (2012) to better understand the method and quality of Catalist's matching procedure. Catalist updates their registration records several times a year from each jurisdiction, which improves the accuracy of the data. Catalist contracts with and collects information from credit card companies, consumer surveys, and government sources to improve its matching capability. Catalist also de-duplicates records by linking records of the same person listed in a state's voter file more than once and runs all records through the Post Office's National Change of Address Registry to

identify movers. This information is helpful for understanding, for instance, whether college students register to vote after they have moved to a college campus or whether they remain registered at their home address.

To identify a student in the snapshot of voting records, the Clearinghouse uses the student's name, date of birth, earliest known home address, and campus address. Catalist collects a broad range of information on each of the individuals in its database. Records submitted for matching against the Catalist national database are examined through a sophisticated set of proprietary algorithms that utilize statistics, fuzzy logic, and matching learning to determine the highest probability matches. (For specific information about Catalist's matching procedures, see Ansolabehere and Hersh, 2012, pp. 443-445.)

Once the two sets of records (Clearinghouse and Catalist) have been merged, each individual record is accompanied by a confidence score reflecting the similarity of the submitted record to the returned match record given all the possible combinations considered by the matching algorithms. The average confidence rating for students in the NSLVE database is 96%.

The process fully protects student privacy. After the matching process is complete, the Clearinghouse removes all student-level identifying information and sends the anonymized data file to IDHE researchers. All student records are linked to a college or university's OPE ID number, an identification assigned by the U.S. Department of Education's Office of Postsecondary Education (OPE) and the privacy rights of students under FERPA are fully protected in this process. At no time does NSLVE receive the names or addresses of individual students. The only entity that knows the identity of individual students is the Clearinghouse, which already knows the students' identities. Catalist personnel do not have access to the merged data, and the company may not and does not collect or store the data elements needed for the matching process. Any information that might allow the re-identification of an individual student (e.g., fewer than ten students in a particular field of study or demographic group) is replaced by the Clearinghouse with an asterisk. Date of birth is replaced with age on the date of election. The records are stored in a secure research drive at Tufts University.

5. Calculating the Institutional Voting Rate

The combined enrollment and voting records are then aggregated to the institution-level and merged with selected data from the U.S. Department of Education's IPEDS. IPEDS consists of nine interrelated survey components that are collected over three collection periods (fall, winter, and spring), either annually or semi-annually depending on the survey component. The completion of IPEDS surveys is mandatory for all institutions that participate in any federal financial assistance program authorized by Title IV of the Higher Education Act of 1965. The IPEDS system collects a variety of institutional characteristics, including type of institution, size, enrollment, financial aid, and other characteristics. Data collected for the 2012, 2014, and 2016 fall semesters were used in this database.

We use IPEDS to account for nonresident aliens (NRAs) at an institution. We first calculate the percentage of NRAs at an institution using IPEDS data. Next, we use this percentage of NRAs from IPEDS and multiply it by the Clearinghouse's number of students at an institution. Using this method, we generate an estimated number of nonresident aliens at an institution. This number is used in our calculations of an institution's voting rate. To calculate an institution's voting rate, we first calculate the number of eligible voters at an institution (number of respondents minus students less than 18 minus number of NRAs at an institution). We then divide the number of students who voted by this number of eligible voters.

Conclusion

Currently, the NSLVE database contains approximately 29 million records for students from over 1,000 institutions for the 2012, 2014, and 2016 U.S. national elections. NSLVE contains about 9 million records per election, and we are already preparing for the 2018 election and beyond. We continue to recruit new colleges and universities into the study. If your campus is interested in joining, please email idhe@tufts.edu to learn more about the study. The NSLVE database allows us to not only research college student political engagement, but also to offer data-driven recommendations to institutions for supporting political involvement among low-propensity voters in order to increase the political mobility of these students. We provide general resources on our website as well as customized services to campuses interested in shaping practices and campus culture to foster active participation in our democracy. Please visit http://activecitizen.tufts.edu/idhe to learn more and to join our email list.

References

Ansolabehere, S., & Hersh, E. (2012). Validation: What big data reveal about survey misreporting and the real electorate. *Political Analysis, 20*(4), 437-459.

Ansolabehere, S., & Hersh, E. (2013). Gender, race, age and voting: A research note. *Politics and Governance, 1*(2), 132-128.

Ashok, V., Feder, D., McGrath, M., & Hersh, E. (2014). Dynamic voting in a dynamic campaign: Three models of early voting. Retrieved from: http://www.eitanhersh.com/uploads/7/9/7/5/7975685/early_vote_v3.4.1.pdf.

Berent, M. K., Krosnick, J. A., & Lupia, A. (2011). The quality of government records and "overestimation" of registration and turnout in surveys: Lessons from the 2008 ANES Panel Study's Registration and Turnout Validation Exercises. Working Paper No. nes012554. Ann Arbor, MI and Palo Alto, CA: Ameri-

can National Election Studies. Retrieved from: http://www.electionstudies.org/resources/papers/nes012554.pdf

Catalist (2017). Catalist. Retrieved from: http://www.catalist.us.

Dynarski, S., Hyman, J., & Schanzenbach, D. W. (2013). Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion. *Journal of Policy Analysis and Management, 32*, 692–717.

Dynarski, S. M., Hemelt, S. W., & Hyman, J. M. (2015). The missing manual: Using National Student Clearinghouse data to track postsecondary outcomes. *Educational Evaluation and Policy Analysis, 37*(1S), 53S-79S.

Integrated Postsecondary Education Data System (IPEDS). Available at https://nces.ed.gov/ipeds/.

National Student Clearinghouse (2017). National Student Clearinghouse. Retrieved from: http://www.studentclearinghouse.org.

About the Authors

Dr. Nancy Thomas directs the Institute for Democracy and Higher Education and is the principal investigator for the Institute's National Study of Learning, Voting, and Engagement. Her work and scholarship center on higher education's democratic mission, college student political learning and engagement, free speech and academic freedom, and deliberative democracy on campuses and in communities. She is the author of multiple book chapters, articles, and the monograph, Educating for Deliberative Democracy, an issue of Jossey Bass' New Directions for Higher Education series. She received her B.A. in government from St. Lawrence University, a J.D. from Case Western Reserve University School of Law, and an Ed.D. from the Harvard Graduate School of Education.

Dr. Jodi Benenson is an Assistant Professor in the School of Public Administration at the University of Nebraska Omaha. Her primary research interests include civic engagement, inequality, social policy, and nonprofit organizations. Previously, Jodi was a Postdoctoral Scholar at the Jonathan M. Tisch College of Civic Life at Tufts University. At Tufts, she worked on several projects in the areas of youth civic engagement, political learning and engagement in higher education, and national service and employment. Jodi received a B.S. and M.P.A. from Indiana University and a Ph.D. in social policy from the Heller School for Social Policy and Management at Brandeis University.

The purpose of the study was to collect data on how the transition to distance learning impacted undergraduate and graduate students taking courses in public health at GSU. The goal was to identify student academic challenges and the unforeseen benefits of distance learning, and to use that information to inform practices that can be implemented during crises that impact university education. Educational engagement can be defined as a state of being when a person is cognitively, behaviourally and emotionally involved in learning activities, and is maintaining a heightened sense of concentration, interest and enjoyment during those activities (Fredricks, Blumenfeld, & Paris, 2004; Gibbs & Poskitt, 2010; Shernoff, 2013). Â Thus, educators work towards providing optimal conditions that maximise the educational engagement and the learning of individual students (Shernoff, 2013). As noted in the introduction, the middle years of schooling (Years 7 to 10) have been identified as a transition period in which disengagement with education is particularly likely (Balfanz et al., 2007; Gibbs & Poskitt, 2010; MCEETYA, 2008). This study provides validation of the Online Student Engagement scale (OSE) by correlating student self-reports of engagement (via the OSE) with tracking data of student behaviors from an online course management system. Â Creating and validating research tools to measure various aspects of the online teaching environment is an important part of advancing research about online learning (Roblyer & Wiencke, 2004). Such tools may also provide feedback to instructors about individual courses. Â learning by taking part and maintaining relations with others. It is a complex process comprising doing, communicating, thinking, feeling and belonging, which occurs both online and offlineâ€ (p. 1761).Â His definition gave rise to the National Survey of Student Engagement (NSSE).