

Florida TQ Data Landscape (K-20 Education Data Warehouse)

The state collects achievement test data for Grades 3-10 from the Florida Comprehensive Achievement Test (FCAT) as well as information on the courses in which students are enrolled, minutes of instruction, and whether or not a student is being mentored. The state has the capacity, because of common course numbers, to link teachers and students. Some policymakers have a strong interest in using student test scores to assess teachers. (A Responsible Teacher File contains the Social Security Number (SSN) for each teacher by grade and subject. This number can be matched with FCAT scores to allow fixed effect models, i.e., value added models.)

In January 2001, Florida began development of its K-20 Education Data Warehouse (EDW). See warehouse schema at <http://edwapp.doe.state.fl.us/doe/EDW.htm>.

Types of Data

The K-20 EDW integrates data from 26 state-level operational source systems and provides a view across systems that reflect the K-20 education environment. The EDW was created to unify data collection efforts and to allow agencies to conduct trend analyses and track students and teachers, to evaluate programs such as Reading First and summer school, and to track course taking of retained ninth graders. Data sources include student background, staff data, and certification and draw on vocational education, adult education, community colleges, and Florida's public universities (including graduate schools). The system has 6-7 years of linked data (by SSN) that allow researchers to follow the individual progression of teachers from high school through public colleges and into teaching, where a teacher can be linked to his or her students and state assessment scores. As data are linked, basic checks on the cleanliness of the data are performed. The EDW also contains basic teacher data including addresses, demographics, and certification data. State agencies are attempting to deal with issue attributing effects to teachers where the student has two teachers of the same subject.

The system began by extracting existing data on individual students and staff going back to 1995-96 as data permitted. The warehouse includes data on courses taken, schools attended, test scores, teachers, and progression through grades. The data on teachers are extracted from the Florida Department of Education databases and cover teacher demographics, salary and compensation, course assignments, licensure test scores, degrees earned, licensure status, and endorsement areas. Warehouse data are longitudinal and are refreshed as they become available from the source systems.

At the warehouse each teacher record is assigned a different unique ID. Each year new assignments of IDs are checked using race/ethnicity, birthday, and name. The database includes both staff data on teacher demographics and current work as well as certification data with information on preparation. The database also has individual records from public higher education institutions that can allow tracking of teacher preparation

graduates to work in the schools including information on their coursework (some analysis has been completed using these data). The data have SSNs, which allows linking with other data in the system. Officials can access a separate file at the Board of Governors that includes majors. They are now in discussion with EDW to provide them with this file, which goes back to 2000; similar files date back to 1992. The data they provide include information on higher education student demographics, course grades, degrees, financial aid, high school SAT score, and high school attended. The primary weaknesses of these files are the lack of a flag for those in teacher preparation who have attended several higher education institutions and lack of data on teachers who are prepared in private institutions or out of state. It is uncertain why data are not assembled on teachers being prepared at private colleges and in the community colleges, which are developing teacher educator centers to help certify more teachers.

Use of Data

The teacher data are used for “highly qualified” teacher designations, Equal Employment Opportunity (EEO) reporting, and responses to legislative requests as well as data reported to the National Center of Education Statistics (NCES) and the National Education Association (NEA). Staffing data drives budget allocations; funding is based on current year staffing, measured three times a year.

The K-20 EDW does not define new data elements on students, but instead uses the existing data and standardizes those data elements as needed. For the match process a separate identification database that contains personally identifiable information was created. During the match process, a randomly generated internal ID is assigned to the records. Thus, all identifying information is kept separate and secure, away from the warehouse database that is available for reporting, research, and program analysis. However, the assigned ID is kept in the warehouse data and may be released to data users. This enables users to add data in the future without being able to identify individual students. Students cannot be identified because the assigned ID is drawn from a list of unique randomly generated numbers. As additional protection of student privacy, tables and reports prepared from EDW data leave cells masked if counts are for fewer than 10 students. Users functioning on their own also must agree to this restriction.

The K-20 EDW adds value to Florida’s existing databases by joining complementary databases that would otherwise be kept separate and by keeping EDW data in longitudinal form. Particularly noteworthy is their data sharing with the Florida Education and Training Placement Information Program (FETPIP), which tracks students from school to work and is updated quarterly. The data may be used to measure student attrition from teacher education programs as well as turnover and “moonlighting” among teachers employed in Florida’s public schools.

Data Requests

In order to guard against inappropriate use of data, potential users are required to submit requests for the specific data they seek. These requests are reviewed by persons who are

familiar with the limitations of the data, and requesters are warned of uses that would go beyond what the data would support. Data access approval takes less than a month. The data warehouse includes a “front end” that produces data in aggregated forms in order to make data releases small enough to be manageable (the size of the warehouse is immense). However, the agency’s capacity to deal with a high volume of public data requests is uncertain, as is whether or not agency staff have experience with a number of potential sensitive data requests.

Cost

The K-20 EDW went into production in 2003 and is just now beginning to crank up to its potential. The warehouse currently employs 12 full-time and 1 part-time technical staff, who maintain and edit data and make them available to users.

The K-20 EDW is owned and operated by the Florida Department of Education, which exercises some governance and oversight responsibility over all levels of education in the state. The cost of building the warehouse, including software, hardware, communications equipment, and consultants, was approximately \$7.1 million during 2001-03. The EDW annual budget is approximately \$2 million, which includes staff, support, and data processing. The warehouse has a separate line in the budget and seems well established in the state.