

NRC Workshop—Completers and Dropouts—October 2008

How can states build and maintain data systems capable of collecting and storing required data to compute these indicators?

Discussant notes—Deborah Newby

EDFacts

EDFacts is an initiative of the US Department of Education—about 15 years in the making, about four years in actual existence—that seeks to consolidate the various program office collections into one collection, and more importantly, to change the way data is used at the department. As we change how we're collecting and storing data, these changes are also happening at the states. Rather than each program office collecting its own data, we're collecting it through one system. We just recently acquired a data warehouse. We're developing an ad hoc reporting environment that pulls data from the warehouse. So there's potential for much greater flexibility in how data is used with the Department. The first year of a data warehouse can be messy—users soon identify business rules that are needed to ensure that only good data get into the warehouse.

Along with all of these changes, we're also instituting a system of data governance—another phenomenon that's happening in the states as well as at the federal level. Although we're changing the way we collect and store data, the program areas that have a vested interest in using the data must remain actively involved in identifying needed data and defining data—not just data that's collected, but also the metadata, which is critical to valid data use. Data elements collected through EDFacts have definitions assigned to them. Sometimes those definitions come from the Department, but there are many instances where state definitions are involved. Some very significant data gets defined by the states—they define their proficiency levels, improvement targets. They've been defining how they calculate their graduation rates, but new regulations being released are likely to require use of the NGA graduation rate. (The NGA rate fits within the NCLB

definition, and it is becoming the rate used by all states due to the Governor's compact.) Metadata provides the context for understanding the data that's collected.

We began collecting data from the states through EDfacts about four years ago, and we're transitioning to full data collection required through EDfacts for SY 08-09. Technically, we're not collecting anything that hasn't been collected through individual program collections for a couple of decades in some cases, but we're collecting it in a new way, and for some data elements, at new levels (such as state, district, or school level). This transition has put a huge burden on many states.

Looking specifically at dropouts, grads, and completers data: The NCES Common Core of Data began collecting graduates and completers and dropouts through EDFacts in 05-06 as a pilot year--states were given an option of submitting through EDFacts or using traditional CCD file submissions.

For the 06-07 school year data, EDFacts was the only option for submitting CCD data. Virtually all states could report state-level grads and completers through EDfacts; about ½ dozen had incomplete LEA data coming in for grads/completers; about 15 states had incomplete school data for grads/completers. Dropout data in Edfacts for 06-07 is slightly sketchier—Previously, CCD focused on dropout data at the LEA level, so we added other levels not previously collected (state and school). In EDFacts, we're still missing dropout data at the state level from about 4 states; about 10 states reported incomplete LEA data; about 17 states submitted incomplete school data.

When we look at the states that are reporting complete data, it doesn't necessarily sync up with whether or not the state has a centralized data system in place. This is true of other EDfacts data files as well. It's become clear that the states' ability to report data to us—and particularly quality data—is not just a function of whether they have a sophisticated data system in place. And not all data systems seem to be of equal quality. Although states are all building student information systems, they are building these systems on 'legacy' silo systems. It's rare for a state to build a system from scratch.

As we talk with these states that have issues reporting data to us, these are some of the things we're hearing are:

State policies on when data gets collected affects what gets collected, how the data is defined, and their ability to report timely data. States may have long timelines for districts to dispute or correct data that they previously submitted, and the state may not want to submit data to the federal level until all disputes are resolved. They are cautious about submitting any data they may be held accountable for.

Another factor that affects state's ability to submit data to the federal level is the processes in place to create data submission files. Some states have their systems programmed to automatically spit out the Edfacts files. Some states are building files by hand. If there is staff turnover, they may be recreating this year after year if no one is keeping records about how it was done. This becomes a huge resource issue.

If the data is not all centrally stored and available to all offices within the state, there need to be processes for getting the data from the program offices. This may involve some technical capacity, but more importantly, it requires interpersonal relationships. People don't always like sharing their data, particularly if they're not sure how it will be used. Related to this, data governance, as mentioned before, is a big issue in the states now—once they have the systems, they need the processes, relationships, clear owners of the data (program offices) need to be responsible for their data. The IT and central data management offices are facilitators and resources to the program offices

Also: Staff resources in both the data office and the program offices affect the capacity for timely and quality reporting. Staff have other competing priorities.

Leadership and political will seem to be the most important factor in whether data get reported out of the state. A state can have the systems, the resources, etc. But someone in

a leadership position needs to say that data reporting is important and a priority, and require the program offices and data offices within the agency to work together.

Responses to presentations:

What we heard from Nancy is that the DQC survey indicates that states ARE building data systems capable of collecting and storing required data to compute graduation and dropout data. 49 states say they have systems to report a cohort graduation rate. But not all of these systems will produce the same quality data. Nancy dug a little deeper into state capacity and looked at their ability to collect exit codes—this impacts the quality of the data they crank out.

Lavan outlined characteristics of a good system—or what a good system produces. Lavan’s description of what was common in the 1980s still exists for some states for some collections. There are still a few states that have paper and pen data collections for some types of data.

We noted the explosion in data that is being collected and maintained by state agencies as they move to individual student records away from point-in-time aggregate data collections. We also highlighted the NCES best practice document: Accounting for Every Student, A Taxonomy for Standard Student Exit Codes. This document is available free from ED pubs. (search at nces.ed.gov) This publication includes the results of a survey of states regarding the exit codes they use. This data is about three years old now, but shows the variations in how states collect data.

Bill went into what’s happening at LEA level. This is where great potential lies for using data to improve program delivery and design intervention strategies. Bill mentioned key data points that they can use to identify students at-risk—most notably, attendance records. Data systems need a system of ‘flags’ in order to identify students that are exhibiting risk factors.

Rob recommended not building data systems just for reporting purposes, but rather to build them for instructional improvement, or other goals identified by states. This is a key rule of data quality: collecting data that is not used by the individual(s) responsible for producing the data is likely to result in bad data. If those gathering the data have a reason to use it, they have a vested interest in making sure it's good data. Rob also reiterated that the code set behind a data system is what helps determine quality and validity.

Which brings us to data use: How is data used? Depends on level. At LEA—ideal use of data is for instructional improvement and designing and implementing effective intervention programs.

At the state level—data historically collected for compliance reporting; now for accountability as well. And states have a new role: assisting districts in need of improvement, so states are getting involved in instructional strategies. Some states also have a role in making data available back to districts, and assisting them with data use. Sophistication of district data systems varies greatly—in some states, districts are head of states, but in those some states, there are districts with very limited resources that need help from states in this area.

At federal level—we're seeing a change in how data used. Not too long ago, data was collected about the program—looked at inputs, how many served. Now looking at outputs—what are results of program—is program administered effectively? Is it making a difference? Where is it making a difference? So program effectiveness is one kind of use at the federal level. Second kind of use is accountability, an outgrowth of the standards-based reform movement. All students should essentially be exposed to same standard of education. Now, we're looking at outcomes for students against those standards. Are outcomes the same for all populations? Civil rights data collection—I consider to be a form of accountability—are all students served appropriately according

to the law? And a third use at the federal level—the data collected is used to create a statistical data set. NCES doesn't just take what they collect and put it out there for use; goes through all kinds of statistical gyrations to make it worthy of statistical analysis. They are the keeper of large, historical statistical data sets for research purposes. But it should be noted that they do not currently collect, report, or make available proficiency or performance data, aside from NAEP data.

Final thoughts:

The workshop discussion has been focusing quite a bit on graduation rates. Real guidance is needed around other types of completer rates and also dropout rates. A system of multiple indicators is needed to get a complete picture of how students are served by our system of education: they may graduate with a traditional, 'on-time' diploma; they may be receive a standard diploma in more than four years; they may earn other types of completion credentials; they may drop out without any kind of credential and not return; they may age-out (no longer eligible for services) without a credential. Graduating or dropping out are not the only two 'final outcomes' for students in our system.