Ensuring High-Quality Data

Transforming SLDS data into meaningful measures can be an enormous undertaking that requires a high level of effort and analytic expertise.

SLDS are based largely on administrative data from state and local sources. Typically, analysts must construct longitudinal, student-level analysis files from annual data submissions obtained from multiple stores within the data warehouse. Analysts may also need to construct some measures of interest from the administrative data elements. This data preparation process can be time consuming and requires careful attention to creating valid and reliable data. The pilot involved consolidating data from 300 million education records (from multiple data sources over multiple academic years in three states) to create 4 million unique student longitudinal records. Ultimately, dozens of measures were constructed from hundreds of source data elements. Detailed materials that describe the pilot’s data approach and key measures are available in the Data Development section of the pilot dissemination web site.

Engaging the ultimate data users in measurement decisions improves the usefulness and trustworthiness of resulting information.

Many decision rules need to be determined when transforming SLDS data for actionable reporting. It is important to seek input from the intended users of the data on key decisions and at each review step, in addition to consulting state-level research and data staff. The involvement of end users will ensure that the resulting data are not only technically reliable, but also valid for the users’ purposes. For example, in one pilot state, missing data on summer graduates made a small difference in the state-level count of high school graduates, but levels of missing data varied widely at district and school levels. It was not until district- and school-level staff reviewed their own data reports that the problem of missing summer graduates surfaced. These users emphasized the importance to them of including outcomes on these graduates in their overall results, so the missing data were obtained and incorporated. Involving local staff earlier in the data preparation and review process could have avoided the discrepancy. Some data problems or decisions may not be apparent in state- or even district-level data and ensuring review at the school level can be important.