Challenges in Matching Student and Worker Information Raise Concerns about Longitudinal Data Systems

Why GAO Did This Study
From fiscal years 2006 through 2013, the Departments of Education and Labor provided over $640 million in grants to states through the SLDS and WDQI grant programs. These grants support states’ efforts to create longitudinal data systems that follow individuals through their education and into the workforce. Analyzing data in these systems may help states improve outcomes for students and workers.

GAO was asked to review the status of grantees’ longitudinal data systems. This report examines (1) the extent to which SLDS and WDQI grantees match individual student and worker records and share data between the education and workforce sectors and (2) how grantees are using longitudinal data to help improve education and workforce outcomes. To answer these questions, GAO analyzed data from a 2013 survey conducted by the DQC. This survey collected information from states on data linkages among education and workforce programs and on how states use longitudinal data. In addition, GAO interviewed a nongeneralizable sample of five grantees, which were selected based on the progress they have made in matching data and on the funding they have received from the SLDS and WDQI programs. GAO also reviewed relevant federal laws and regulations.

GAO is not making recommendations in this report. GAO received technical comments on a draft of this report from the Department of Education and the Department of Labor, and incorporated them as appropriate.

What GAO Found
Over half of 48 grantee states that received a Statewide Longitudinal Data Systems (SLDS) or Workforce Data Quality Initiative (WDQI) grant have the ability to match data on individuals from early education into the workforce, based on GAO’s analysis of 2013 Data Quality Campaign (DQC) survey data. The DQC is a nonprofit organization that supports the effective use of data to improve student achievement. In its survey, DQC collected self-reported information from states on their ability to match, or connect the same individual record, between the (1) K-12 and early education, postsecondary, and workforce sectors and between the (2) postsecondary and workforce sectors. However, as the match rate—that is, the percent of unique individual records reliably connected between databases-increases, the number of grantees able to match data decreases. GAO found that more grantees reported being able to match data among the education sectors than between the education and workforce sectors. Further, most grantees reported that they are not able to match data comprehensively. For example, only 6 of 31 grantees reported that they match K-12 data to all seven possible workforce programs covered by the DQC survey, which include adult basic and secondary education as well as unemployment insurance wage records. State officials cited several challenges to matching data, including state restrictions on the use of a Social Security number. Specifically, officials in three of five grantee states GAO spoke with said state law or agency policy prohibit collecting a Social Security number in K-12 data, which can make it more difficult to directly match individuals’ K-12 and workforce records.

According to GAO analysis of the DQC survey data, grantees use some longitudinal data to inform policy decisions and to shape research agendas. All 48 grantees reported analyzing aggregate-level data to help guide school-, district-, and state-level improvement efforts. For example, 27 grantees said they analyze data on college and career readiness to help schools determine whether students are on track for success in college or in the workforce. Grantees also reported using longitudinal data to analyze outcomes for individual students. For example, 29 grantees reported that they produce early warning reports that identify students who are most likely to be at risk of academic failure or dropping out of school. Data from the DQC survey also show that 39 grantees reported developing a research agenda in conjunction with their longitudinal data systems.