Data Access, Data Sharing, and Data Use in Education Research—Advancing Knowledge through Responsible Conduct

Felice J. Levine
American Educational Research Association

OECD Workshop on Fostering Innovation in Education: The Contribution of Longitudinal Information Systems

June 30 – July 1, 2014
New York, NY
Key Questions in Responsible Access and Use

• Why share education data and information in (longitudinal) systems of data?
• Who are the customers for country, state/province-level statistics and data agencies/units?
• What are the forms of data and information being collected by education institutions and agencies?
• What is the relevance of consent, privacy, confidentiality, and data protection in this context?
• How can data best be shared and used? What are the issues involved in responsible access and use?
Why Share Education Data and Information in (Longitudinal) Systems of Data?

- Expands Breadth of Usable Knowledge
- Increases Use and Range of Uses—Designed and Emergent
- Allows for Verification, Replication, Testing of Alternative Questions/Rival Hypotheses; Builds Cumulative Knowledge through Longitudinal Data Over Persons and Over Time
- Transforms a “Warehouse for Storage” to a “Warehouse with a Moving Inventory”
- Improves Quality of Measures/Indicators/Observations and Overall Systems of Data
- Encourages Investment in Data as a “Valued” Asset
Who are the Customers for Country, State/Province-Level Statistics and Data Agencies/Units?

• Agencies Themselves Producing Data, Reports, and Products that Educational Institutions Cannot Produce as Readily (or with same level of expertise or independence in data evaluation, analysis, and reporting)

• Education Researchers—Sometimes in Partnership with School Systems or Agencies—as Knowledge Producers and Expert Sources for Longitudinal Data System Improvements

• Educational Institutions (P-20) With Access for Internal Use

• Students, Families, Teachers/Faculty, and Educational Leaders Learning Short- and Long-Term about Elements of Effective Education and Peer Comparative Data

• Public Users (including Media and Policy Groups)
What are the Forms of Data and Information Being Collected by Education Institutions and Agencies?

• Administrative Records on Students, Teachers/Faculty, and Institutions (non-research records, including personal private information); in U.S., increasing use of common data elements and standards in recordkeeping and reporting. See CEDS, v4, NCES, 2014.

• Studies/Surveys/Observations Linked or Complementary to Administrative Records (research data)

• Linked Data from Other Institutions/Agencies—e.g., Health, Justice System, Employment (administrative records)

• Harvested or Gathered Big Data—GIS, Digital Internet Data, Biomarkers, Bio/Psycho-Social/Time-Use Monitors with Academic and Non-Academic Outcomes (such data may be publicly available information or may be generated as part of research)
What is the Relevance of Privacy, Consent, Confidentiality, and Data Protection in this Context?

Thanks to U.S. National Research Council Common Rule Report, 2014

- **Personal Private Information** (PPI), Reasonable Expectation that Such Information Would Not be Generally Known (i.e., beyond name and personal information, if information linked to identifier or name)

- Data Systems with Identifiers, But Where Individuals Have No Reasonable Expectation of Privacy are not PPI Data (e.g., court records, telephone directory); **Professional Norms Should Prevail**

- Research Using Public Information, or Observation With No Expectation of Privacy; No Consent Needed; **Professional Norms Should Prevail**

- Public Use Data Sets De-identified for Public Use; Use Responsibly

- Administrative Records with PPI Data Can Provide for Use Without Further Consent Consonant with Privacy Acts (e.g., in U.S., FERPA); But Must Adhere to Confidentiality Agreements and Data Protection Plans as a Condition of Access; Includes Using Linked Data

- Data Collection with Interaction or Intervention with Persons **Requires Informed Consent Agreement** (To be Buttressed by Confidentiality and Data Protection Provisions); Includes Using Linked Data
Appendix: Beyond FERPA—State Responsibilities and Critical Questions to Protect Student Data

Through a **common understanding of and commitment to privacy and security principles, addressing legal roadblocks** preventing appropriate data use, and **providing sensible implementation and oversight of strong policies and practices** that protect student data from harm, the education sector can maximize investments in data systems, minimize data risks, improve data quality, and increase data management efficiency.

State policymakers have three overarching responsibilities to help protect the privacy, security, and confidentiality of students’ personally identifiable information.

**Establish roles for data stewardship:**
Define and clearly communicate authority, responsibility, and accountability for decisionmaking, management, and security of data.

**Ensure policy documentation, transparency, and enforcement:** Document laws, policies, and decisions related to data governance and communicate these policies and procedures in a way that is accessible to stakeholders, including agency staff, students, parents, and the public.

**Support organizational capacity:**
Ensure the state has the capacity and resources to implement and sustain these policies and procedures, including staff and technical system infrastructure.
**Examples of Data Provisions and Use**

Excerpted from DQC March 2013

<table>
<thead>
<tr>
<th>Applying FERPA to Common Policy Scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is critical for policymakers to understand how FERPA applies to common scenarios related to high-priority efforts in their states. See below for common scenarios based on information drawn from regulations and ED's materials.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STATE LONGITUDINAL DATA SYSTEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does FERPA permit schools and local education agencies, without parental consent, to provide students' education records to a state longitudinal data system?</td>
</tr>
<tr>
<td>Schools and local education agencies may disclose student records to a state education agency or other centralized state data entity, such as a statewide longitudinal data system, for purposes of evaluating (or auditing or ensuring compliance of) federal or state-supported education programs.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIGH SCHOOL FEEDBACK REPORTS OR COLLEGE SUCCESS INDICATORS BY HIGH SCHOOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does FERPA permit postsecondary entities or agencies to share student-level postsecondary data with the state education agency or districts for purposes of calculating postsecondary student enrollment and remediation rates by high school or school district?</td>
</tr>
<tr>
<td>The 2011 regulations allow such disclosures for purposes of evaluating how well districts and public schools prepared students for college.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SHARING DATA ACROSS STATE LINES</th>
</tr>
</thead>
<tbody>
<tr>
<td>May individual, student-level data be shared across state lines?</td>
</tr>
<tr>
<td>The preamble of the 2011 regulations states that nothing in FERPA specifically prohibits interstate disclosures that are made for the purposes of, and are consistent with the requirements of, the regulations.</td>
</tr>
</tbody>
</table>
How can Data Best be Shared and Used? What are the Issues Involved in Responsible Access and Use?

- Public Use Files for Longitudinal Data and Data Systems De-identified for Research and Other Use
- Use of Individual-Level Data Under Restricted Conditions
  - Research Data Centers
  - Third Party Repositories
  - Site Licenses to Researchers through Their Institutions
  - Secure Access—Researchers submit requests for analyses; provided by data stewards
- Assess Disclosure Risk, Statistical Methods Also Available
  - Data Alteration—Swapping, Masking, Attenuating Tails
  - Data Simulation
- Emphasis Should be on Data Dissemination and Reporting; Test and Protect Against Deductive Disclosure
Take Away and Next Steps in Responsible Data Access and Use

• Develop Workshop Recommendation that Reflects Joint Commitment of Data Agencies/Providers and Research Users to Access and Use
• Foster Culture of Data Access and Use in “Own Backyard” and By Researchers with Access to Individual-Level Data
• Provide Information and Meaningful Communication to Other Stakeholders—Families, Teachers, Education Leaderships, Policymakers, Media on Ethical Principles and Practices in Data Production, Sharing, and Use