Institute of Educational Statistics National Center for Education Statistics

## NATIONAL INSTITUTE OF STATISTICAL SCIENCES EXPERT PANEL REPORT

## SETTING PRIORITIES FOR FEDERAL DATA ACCESS TO EXPAND THE CONTEXT FOR EDUCATION DATA

## TABLE OF CONTENTS

Exect	utive	Summary	3			
Prefa	ice		6			
I.		BACKGROUND				
II.		OVERVIEW	8			
III		TOPICS, DATA AND LINKAGES	8			
	Торі	CS	9			
	Data	۱1	0			
	Linka	ages1	0			
IV	•	IMPLEMENTATION1	1			
V.		PRIORITIZATION1	1			
	Purp	oose1	2			
	Prine	ciples1	2			
	Proc		2			
	Start	ting Points1	3			
VI		GOING BEYOND	4			
VI	I.	SUMMARY OF FINDINGS	4			
	Recommendations15					
AP	PEN	DICES1	9			
	Appendix A: Agenda20					
	Appendix B: Charge to Panel21					
	Appendix C: Case Studies22					
	Appendix D: Use Case Scenarios					
	Domain: Criminal Justice and Safety					
	Domain: Environmental (Physical/Geographical) Influences					
	Domain: Non-traditional Education and Non-degree Credentials for Work					
	Domain: Education Careers and Labor Markets					
	Dom	nain: Effects of Disruptions4	1			
	Appendix E: Expert Panel Biosketches43					

## NATIONAL INSTITUTE OF STATISTICAL SCIENCES

# SETTING PRIORITIES FOR FEDERAL DATA ACCESS TO EXPAND THE CONTEXT FOR EDUCATION DATA

## EXECUTIVE SUMMARY

Passage of the Evidence Act in July 2019 opened opportunities to integrate administrative data across agencies. For the National Center for Education Statistics (NCES), this opportunity to broaden the context for NCES data bases expands the potential scope for education research and enables enrichment of the information base for educational policy. In 2021, NCES charged the National Institute of Statistical Sciences (NISS) with convening a panel of experts to consider the potential benefits from linking information from other federal agencies with NCES data bases and to offer suggestions for establishing priorities for selecting data bases to be linked or integrated.

Specifically, the panel was asked to set priorities for accessing federal data outside the Department of Education, to identify the potential value and the challenges in establishing these links. The goals for enabling this enrichment of NCES data are to enhance this primary education information resource and also to support external objectives that require education data. In responding to this task, the panel identified a number of domains of national interest to serve as examples and as a basis for focusing their deliberations and formulating recommendations. Several of these are included in the appendix to the full report.

The starting point for panel discussions was ultimately reiterated in their findings and conclusions.

- Meeting the information needs of NCES stakeholders is critical to NCES's continuing relevance as a data source for researchers, policy makers, and IES/Department of Education.
- Many critical issues involving education and education information can only be informed by integration of data from other sources with NCES data.
- Therefore, expanding the context for IES/NCES education data that is available to IES/NCES stakeholders is a matter of some urgency not to be relegated to a wish list.

The panel's findings addressed determination of priorities, development of a process to guide data integration or file linkage, and investment of needed resources. Key findings follow in abbreviated form.

## I

#### Focus on Key Questions for Addressing High Priority Issues

• **Provide** a solid information basis for addressing key contemporary issues for decisions by policy makers and other stakeholders, for advances in research directions, and for internal use by the Department of Education. This priority derives from the *importance of the questions* that require integration of data from external sources with NCES data.

• **Define** *importance* in terms of priorities and needs for information of the user community: Researchers, Policy Makers, Educators, Department of Education.

## Expand the Context for NCES Studies and Data Resources

- **Create** a file or set of files of contextual information as a "backbone", to be consistent over time and content, to link to multiple surveys and other data collections.
- **Recognize** the unique value of NCES longitudinal studies, in particular to establish an extensive core of information at both aggregate and micro-levels.

## Fill the Important Gaps in Information

- Access external sources of information on current education preferences and practices to expand access to contextual information beyond CCD and IPEDS<sup>1</sup>.
- Identify external sources with evidence of shifts in education preferences and practices, in particular alternatives and non-traditional options, population and teacher mobility, and large-scale disruptions, of which COVID-19 presents a prime example.

## Π

## Define a Process for Prioritization

- **Establish** guidelines and criteria for requests for data linkage/integration encompassing the rationale, documentation requirements and expectations for release for use by other researchers.
- **Balance** issue importance with effort/resources/time. Consider both the impact of decisions dependent upon the linked files and the breadth of user community able to access these data. Add to development costs the continuing costs of user support, curation and maintenance of files and tools.
- Attend to privacy and data confidentiality, including not only individual person information but also small aggregates or highly cross-referenced data. This is a critical and increasingly complex item as technology continues to advance.

## Maximize Effectiveness of Existing Research Resources

- **Promote** use of EDGE and other NCES tools by increasing visibility and accessibility on the website and expanding available tools and user support.
- **Improve** the website for locating research resources; rethink and restructure the website to advertise and provide direct access to these resources including tools, data collections and publications.
- **Facilitate** creation and sharing of linked/expanded NCES files by partnering within the Department of Education, with other federal agencies and with states.

<sup>&</sup>lt;sup>1</sup> CCD: Common Core of Data on public elementary and secondary schools (at NCES); IPEDS: Integrated Post-secondary Education Data System at (NCES)

## III

## Invest the Resources in Expertise and Effort Required.

- **Prefer,** as a general practice, tools for linkage over single purpose data sets.
- **Expand** and strengthen geo-based tools, for linking data from multiple sources, federal and non-federal.
- Address the need for sufficiently increased effort and technical skills to implement data integration and linkages with cartographer-statistician and computer scientist-statistician teams.
- **Develop** and enhance long-term inter-agency relationships around high priority common interests to expand the scope for of shareable information, micro-data and data-linkage and analysis tools
- Eliminate or work to overcome systemic hurdles, both technical and operational.

## NATIONAL INSTITUTE OF STATISTICAL SCIENCES EXPERT PANEL REPORT

## PREFACE

## Setting Priorities for Federal Data Access to Expand the Context for Education Data

Passage of the Evidence Act in July 2019 opened opportunities to integrate administrative data across agencies. For the National Center for Education Statistics (NCES), the Evidence Act this opportunity to broaden the context for NCES data bases expands the potential scope for education research and enables enrichment of the information base for educational policy. In 2021, NCES charged the National Institute of Statistical Sciences (NISS) with convening a panel of experts to consider the potential benefits from linking information from other federal agencies with NCES data bases and to offer suggestions for establishing priorities for selecting specific data bases to be linked or integrated.

The panel was specifically asked to set priorities for accessing federal data outside the Department of Education, to identify the potential value and the challenges with the goals of enhancing primary education information and supporting external objectives that require education data.

The panel of experts held four virtual meetings between 17 February and 10 March 2021. At the first meeting NCES staff presented details of the panel's charge. Following deliberations in closed sessions, the panel presented preliminary findings to NCES staff at the fourth meeting. Further panel teleconferences were held during the preparation of this report.

## NATIONAL INSTITUTE OF STATISTICAL SCIENCES EXPERT PANEL REPORT

## SETTING PRIORITIES FOR FEDERAL DATA ACCESS TO EXPAND THE CONTEXT FOR EDUCATION DATA

## I. BACKGROUND

The data collections of the National Center for Education Statistics (NCES) are highly valued and trusted for their depiction of education in the United States, for their relevance to the education enterprise, for their national scope and especially for the high quality of the data.

The community of users of NCES data is extremely broad, from individual participants (students, parents, teachers) to administrators and education policy makers at every level from school to federal government. In addition, these resources are needed outside the education enterprise, for example by health care/epidemiology officials and researchers, by urban planners, and by fiscal offices and legislators who decide permits and restrictions, allocation of funds and resources for other enterprises that directly or indirectly impact the education enterprise.

The range of uses of NCES information is equally wide, with impacts that range from individual education option decisions to national policy. The scope of questions often extends beyond what is in the NCES data files; so answers or the available detail and/or local relevance of answers is limited. Yet the needed additional information may be available in data files of other federal agencies. Questions about school conditions, for example, may be broadly answerable on a national or state scale from NCES data. However, for a school or for a school district, information for immediate environment could reside in socio-economic, health, or environmental information federal data bases outside NCES. Passage of the Evidence Act in July 2019 emphasized the importance to the country of learning from empirical evidence. A key to this is integrating data bases from NCES and other federal agencies, and where possible allowing for linkages with data sets collected by states, localities and other entities. The goal of data linkage is to find efficient ways to jointly access data elements from another federal agency data file with the NCES data relevant to answering the questions posed.

This potential could expand the scope of education research and create a deep contextual base of information for education policy. At the same time, data users and researchers outside the education domain would have access to NCES data integrated with information on labor force, facilities, location-specific neighborhood data and demographics. Further, respondent burden may be decreased when auxiliary federal data bases can provide the information at a suitably specific, but not necessarily individual, level.

In considering how to move forward following the Evidence Act, NCES charged the National Institute of Statistical Sciences (NISS) with convening a panel of experts. The charge to this panel was to consider the potential benefits from linking information from other federal agencies with NCES data bases and to offer suggestions for establishing priorities for selecting data bases to be linked or integrated. The panel was

composed in halves of researchers from outside the federal sector and of federal government staff engaged with national statistical data resources.

## II. OVERVIEW

The panel approached the challenge of setting priorities for accessing federal data outside the Department of Education by first articulating the goals for this data integration, then addressing the potential value and the challenges in establishing these links.

In responding to this task, the panel identified a number of domains of national interest to serve as examples<sup>2</sup> and as a basis for focusing their deliberations and formulating recommendations. Several of these are included in the appendix to this report.

The goals for the enrichment of NCES data by integrating or linking to other federal data are to enhance information on education for research and other education purposes, and in addition to support external objectives that require education data.

The premises that the panel took as the starting point for its discussions were:

- Meeting the information needs of NCES stakeholders is critical to NCES's continuing relevance as a data source for researchers, policy makers, and IES/Department of Education.
- Many critical issues involving education and education information can only be informed by integration of data from other sources with NCES data.
- A key role that NCES plays in making data available is vetting the data so that users of both the data and the research from the data trust the source.
- Therefore, expanding the context for IES/NCES education data that is available to IES/NCES stakeholders is a matter of some urgency.

The panel's goals were to address in a useful way: determination of priorities, development of a process to guide data integration or file linkage, and investment of the needed resources, supplemented by examples and illustrative use cases.

The next two sections of this report give details of the panel's discussion of elements that determine value for linkages and some of the requirements for making linkage feasible and practical. The remainder of the report focuses on prioritization and recommendations for moving forward.

## III. TOPICS, DATA AND LINKAGES

NCES data collections cover various aspects of education, making linkage to an equal variety of domains of federal data valuable for both contextual and outcome information. The disruption due to COVID, because of the national scale and the permeation of every facet of life in the US, put into sharp relief relationships of the changes occurring in many domains with education.

<sup>&</sup>lt;sup>2</sup> NOTE: This list is <u>not</u> intended to reflect an ordering of any sort. Rather these examples served to keep discussions concrete rather than purely hypothetical and/or vague, and therefore were selected by panelists to match their expertise.

#### Figure: Paradigm



In particular, the COVID disruption of the education process and practices has led to new questions and solutions for school adaptations to virtual learning, resource allocations to equip school facilities, and decisions about timing and risk-assessment. Also questions about career decisions by teachers and learning options for students – Pre-K, K-12, post-secondary via traditional or non-traditional choices, hiatus, drop-out or adult education.

Relevant contextual information includes SES, income, housing and employment plus health and disability information at the personal level. Environmental information includes geography, hazards, neighborhood demographics, resources such as WiFi, and safety/crime. In addition to measures focused on education choices and success, important outcome measures are economic and labor/employment history, long-term health and mental health, school safety measures, personal justice history.

#### TOPICS

Topics to consider for expanding the context for education data are of two kinds. First, there is broadly applicable basic information, such as might be included in an omnibus survey (like the American Community Survey). Second are the anticipated information demands for policy and research on the anticipated "big issues" of the decade.

The panel compiled a list – in no way complete – of some topics of the latter kind and selected among them to focus thinking in concrete terms and to create working examples of use cases (marked \*) and two case studies (marked \*\*).

- Digital learning environment
- Non-traditional education Learning options for K-12
- Post-pandemic playout (first and second-order effects) \*
- Non-traditional education Non-degree credentials and return on investment \*
- Long-term life paths and outcomes
- Post-secondary education and marketplace/tax/poverty \*\*
- Education careers and labor markets \*
- External social influences and intervention impacts
- Child health and welfare Role of schools and impact on education
- Justice, crime and school safety \*
- Social equity and equity in education
- Environmental (physical/geographical) influences \*
- Geographic area/geospatial data \*\*

## DATA

Data can be generally classified according to granularity – either unit/micro-data, or aggregated at some level. Data can also be classified by type – original, perturbed or semi-synthetic, imputed, or fully synthetic or simulated. Taken together the granularity and type of data determine the accuracy and precision of the linkage, the linkage feasibility, and the extent of restrictions on access and use that are required to preserve privacy and confidentiality.

Aggregate information, such as small group statistics or small-area estimates, is usually easier to obtain, less difficult to match to NCES data and least restricted for research use. Hence, whether defined by geography or by demographics, small-scale aggregate statistics are preferred unless the additional accuracy and/or precision justifies linking micro-data with NCES.

While, at least in concept, original data might be desired, and it is sometimes used for establishments (e.g., schools), data for individuals is regularly perturbed or semi-synthetic (altered by random error, data swapping or other techniques to obscure identity). Especially for contextual variables, refined estimates for small-scale aggregates are often available via imputation, small area estimation or small group pooling can be used quite satisfactorily for most inferences. As a practical matter, these small-scale imputed values are better refined than the values from broader regions or less specifically defined groups that are now frequently cited.

The case for fully synthetic data can be made for entire research data sets that supplant actual data where original data cannot be made available for whatever reason. Depending on the sophistication of the data synthesis and the resources invested in its creation, such a data set can provide a good basis for research investigations. However, it is not clear that these have any role as additions to data in the NCES collections.

## LINKAGES

Linkage can take the form of data integrated into an NCES data file, or a separate new data file constructed to be accessed in conjunction with NCES data collections, or a tool for directly extracting data from another agency's file that uses geography (or demographics) to index data. Each of these is appropriate in circumstances determined by the kinds of data of the NCES collection and of the external data source, the desired granularity/precision and accuracy for linked data, the difficulty of (external) data access, and any restrictions to be imposed on eventual research access and use.

A broad-based file of local contextual information, serving as a "backbone" to support many NCES data collections is an example of the second form of linked data. The idea is to build a solid, comprehensive core of contextual information to be maintained long-term. For resources and omnibus surveys such as ACS that are also geographically indexed, small area aggregate information for neighborhoods, zip codes, or other area unit used by the resource or survey can be linked to school or residence locations.

Tools like EDGE (Education Demographics and Geographic Estimates)<sup>3</sup> which was created by NCES and is available on the NCES website have the most flexible application. Current use of EDGE is extensive; enhancing its capabilities to interface with more data bases that can link via geography will further increase its value. Other federal agencies are also developing and/or contracting development of similar tools for accessing geographically-indexed data. Sharing progress and software may enable the

<sup>&</sup>lt;sup>3</sup> <u>https://nces.ed.gov/programs/edge/</u>

improvement and extension of EDGE and also the creation of new, more universal and more powerful tools.

## **LONGITUDINAL STUDIES**

Longitudinal studies are unique in that they are very high value studies designed to allow tracking individuals' decisions and outcomes. Consequently, they provide some important exceptions to the practice of associating small-area statistics into individual records. In particular for a longitudinal study, data integration is needed at the individual level when either the outcome (e.g., earnings or employment) or the contextual (e.g., enrollment in government program) variable is personal. Here, the substitution of a value for an aggregate would defeat the longitudinal purpose of the study. When federal administrative or universe data (e.g., IRS individual record) allows record-level integration, there is a simultaneous benefit of reducing respondent burden, by pre-populating the data collection instrument. Privacy considerations may require tiered access to selectively tighten restrictions on parts of the integrated data file.

When none of the outcome or contextual variables is personal, even though individual record-level might still be preferred, in most instances (except for universe data) it will not be available. Small-area statistics from other federal data files can still provide local/regional contextual information.

## IV. IMPLEMENTATION

Implementation of a linkage of data from other federal agencies' files with NCES data requires investment of effort, expertise, funding/support, and time appropriate to the task. This is true whether the linkage results in data integration, construction of a new data resource or development of a new data tool.

An intrinsic problem with linkage is the asynchrony between the data resources. When linkage relies on geography, this is manifest in different sizes, shapes and orientations of the geographic units. When linkage relies on attributes, factors or demographics, the differences in definitions, ranges and boundary values must be resolved. With each linkage project, the task is time-consuming and usually requires a high level of technical skill.

When possible, a universal solution to an asynchrony problem is preferable. For example, school district boundaries as geographic units are essential but unique to the Department of Education. Therefore it would be valuable to develop a tool to convert between school district boundaries and the most commonly encountered geographic units such as census tract, zip code, county, etc. Since boundaries crisscross rather than nest, statistical algorithms are needed to take data from all units that overlap partially or wholly with a school district to create an imputed value for that school district, i.e., a small area statistic. The technical approach is the same when the linkage is based on factors or demographics.

It goes without saying that detailed documentation is essential, both for the linkage plan and algorithms and for the code itself to prevent obsolescence with the first change in definition or geographic unit size or boundary. Documented data files are also needed as guidance for users about the data provided at different levels of granularity (e.g., different geographic units), in the correspondence between the external data and the NCES survey frames, and hence, how to link data from other federal source files with NCES data.

Planning for implementation goes beyond the issues in linking data elements and files. Continuing resource requirements may be anticipated depending on expectations for updating of files and maintenance of tools.

Ownership of data, once combined by linkage into a file or as extracted and created by a linkage tool, needs to be determined followed by security requirements and questions of access/license restriction. Tiered access may be the key to meeting the NCES goal of making as much information as broadly available to researchers and policy makers as possible.

## V. PRIORITIZATION

Setting priorities is a combination of taking the long view and taking advantage of immediately available broad-based information and tools. No linkage project is small; and in each case, success depends on the investment of resources and designation of responsibility.

## PURPOSE

The principal goal is to provide a solid information basis for addressing key contemporary issues and for advancing education research.

The user community for NCES data is broad, including policy makers, educators, researchers, and the Department of Education, who hold conflicting and competing interests.

#### PRINCIPLES

Priority derives both from the importance of the questions that can be answered directly and from the breadth of usefulness of the information.

Importance is defined in terms of the priorities and needs for information of the user communities and the potential impact of decisions taken based on this information.

Priority also must balance potential impact against opportunities lost, based on what cannot also be done.

## PROCESS

The goal is to set up a formal process that is as simple as possible.

The first step in defining a process for evaluating priority and recommending specific linkage projects is to assign responsibility for the process. Whether by constituting a decision-making group or appointing an individual, the process should be formalized so that proposals are evaluated according to clear criteria and so that recommendations are clearly based on expected value weighed against opportunities lost.

Criteria are simple to state: the impact of adding contextual data depends on importance of the questions that can then be answered and the breadth of the usefulness of the information to the user community both immediately and in the long term.

Comparative value for creating a linkage depends on the costs of the development and continuing costs, and also on the merits and costs of declining other proposals that compete for the same resources.

The first criterion of importance as a reliable information basis for decisions of high impact relies on understanding the contemporary Big Issues, forecasting those emerging over the next 2, 5 and 10 years, and identifying the critical information gaps. A contemporary example is the changes that will be seen as the nation emerges from the pandemic and schools move away from remote learning as the primary mode.

The second criterion of importance is extent of usefulness. One example is the importance of a backbone of contextual information to be maintained and updated over time with consistent content, comes from the breadth of use. Creating such a cross-sectional and longitudinal array of economic, labor force, and neighborhood information, expands the contextual base for multiple surveys and other data collections.

The third criterion is assessment of the costs of development and of continuing maintenance and updating. These costs begin with the effort to gain access to files followed by extracting, organizing, reframing and otherwise manipulating data to enable integration. Efforts require expertise, personnel, and time for completion. The ancillary results of these efforts may also be to decline or defer other efforts with their own timeframes for usefulness.

With the procedure in place and criteria stated, a framework can be structured for proposing data linkages. The elements of such a framework would need to include:

- Justification in terms of importance, short and/or long term
- Basis for specific files and content to be linked
- Determination of feasibility/difficulty with regard to data access or content integration
- Rationale for linkage type, feasibility and any technical issues or problems to be solved
- Implementation cost/time requirements, proposed resources for development
- Plans for use and access including restrictions, security requirement, continuing resource requirements

The appendix includes several examples (some partial) of hypothetical use cases for issues anticipated to be crucial relative to the field of education over the course of the next decade.

## **STARTING POINTS**

NCES is already engaged in making some geographically-indexed external federal data accessible, and looks to expand this capability. An immediate step is to consider how to exploit existing resources, in particular expand linkages to data held by NCES, Census and others.

Almost as immediately NCES can start to formulate a solid, comprehensive core of contextual information to be maintained long-term that can draw on omnibus surveys and other fairly comprehensive federal data files of economic, labor, health/welfare, federal assistance program (SNAP) and justice data.

Two tasks are crucial to the long-term success of these efforts. The first is setting documentation standards (or minimal essential guidelines) at two levels: i) use and description of all files and of all tools and ii) of code used in defining algorithms to convert between basic unit descriptions (geographic area or demographic data definition) and of all software code used in building a tool.

The second task is to set policies and requirements for access to data files and tools. This applies especially to contractor or researcher-developed files and tools with regard to making these available to other data users.

Longer term benefits will come from developing new inter-agency relationships for sharing microdata and small area statistics, also including agencies outside the federal statistical agencies.

## VI. GOING BEYOND

The nation-wide disruption of education by the COVID pandemic has highlighted the need for information currently lacking in NCES data that exists at local, district and state levels. To understand the consequences of changes forced by COVID, the decisions reached to manage education through changing restrictions and the eventual education impacts, these local/district/state level data will be necessary. These data include operational information such as changes in local practices, attendance information, teacher engagement and retention during and post-COVID.

The need for these data at the level of small-group aggregate is not new – disturbances like Katrina have also disrupted schools and dislocated people. Teacher mobility, especially across states, from graduation to certification to employment and personal relocations, can also be studied from data at these levels.

Previously it has not been regularly or easily possible to bring data from the state or more local levels into NCES data. Since the stakes are high for gaining access to state data, two possibilities are worth investigating. The first derives from the shift in data storage to the cloud. States are able to and do share data with each other via cloud access. Thus sharing does not alter ownership of the data; and (unit level) micro-data can be aggregated before statistics are downloaded. The second is the use of secure computation methodology to prevent visibility while giving direct access to unit level data in the process of creating summary statistics.<sup>4</sup>

Not unique to the disruption by COVID are the unusual questions that overlap other kinds of information held by federal agencies with education in unpredictable ways. To illustrate, consider a hypothetical issue related to school facilities that could draw on data on local housing conditions and occupancy. A HUD database holding relevant information, if geographically indexed, could be linked by an omnibus tool. Such a request by a researcher or a policy maker might not rise to the priority to allocate NCES effort or expertise. But with an available tool to reduce the cost of the project, the resources required for implementation might be available from an external source (public-private/foundation grant), or through NISS partnership with HUD or another agency.

## VII. SUMMARY OF FINDINGS

Three important premises underly the panel's conclusions and recommendations.

- Meeting the information needs of NCES stakeholders is critical to NCES's continuing relevance as a data source for researchers, policy makers, IES/Department of Education.
- Many critical issues involving education and education information can only be informed by integration of data from other sources with NCES data.
- A key role that NCES plays in making data available is vetting the data so that users of both the data and the research from data trust the source.

Therefore expanding the context for IES/NCES education data that is available to IES/NCES stakeholders is a matter of some urgency – not to be relegated to a wish list.

<sup>&</sup>lt;sup>4</sup> Alan Karr (NISS): Karr, A. F., J. Ghosh, and J. P. Reiter. "<u>Secure computation with horizontally partitioned data using adaptive regression splines</u>." *Journal of Computational Statistics and Data Analysis* 51 (2007): 5813-5820. Karr, A. F., S. K. Kinney, and J. F. Gonzalez, Jr. "<u>Data confidentiality—the</u> <u>next five years: Summary and guide to papers</u>." *Journal of Privacy and Confidentiality* 1 (2010): 125-134. Karr, A. F., J. Ghosh, and J. P. Reiter.

## RECOMMENDATIONS

## I

Priority comes in three guises for NCES, not as a hierarchy: issues, foundation and scope.

## Focus on Key Questions for Addressing High Priority Issues

- **Provide** a solid information basis for addressing empirical questions that inform decision-making, for advances in research directions, and for internal use by the Department of Education. This priority derives from the *importance of the questions* that require integration of information from external data with NCES data.
- **Define** *importance* in terms of user community priorities and needs for information; gather information directly from: Policy makers and educators at the federal, state and local levels, Researchers, and the Department of Education.

## Expand the Context for NCES Studies and Data Resources

• **Create** a file or set of files of contextual information as a "backbone", to be consistent over time and content, to link to multiple surveys and other data collections.

For Early Childhood and K-12, information would describe the school district, the school environment, and school neighborhood with added information relevant to the student population, including demographic and socio-economic data. (Expand with new information as warranted and possible.) Outside of EC/K-12, create an equivalent backbone according to the same principles for other data collections.

• **Recognize** the unique value of NCES longitudinal studies, in particular to establish an extensive core of information at both aggregate and micro-levels, integrating both additional context and outcome information.

Prioritize accessing micro-data such as wage, earnings, employment, student loan, occupation, mobility, and social indicators.

#### Fill the Important Gaps in Information

• Access external sources of information on current education preferences and practices to expand access to contextual information beyond CCD and IPEDS.

These gaps include, for example: Local information on practices (e.g., attendance, teacher salaries, budget allocations to technology); Physical plant and facilities; School safety measures and breaches. While this information is held as micro-data at the state/district/local level (not available to NCES) – only aggregate data is needed to integrate or link with NCES files. This could be done under ESRA §9547 Cooperative education statistics systems.<sup>5</sup> (NCES would coordinate activities but not own the data.)

<sup>&</sup>lt;sup>5</sup> "The Statistics Center may establish one or more national cooperative education statistics systems for the purpose of producing and maintaining, with the cooperation of the States, comparable and uniform information and data on early childhood education, elementary and secondary education, postsecondary education, adult education, and libraries, that are useful for policymaking at the Federal, State, and Local levels." From ESRA §9547

• **Identify** external sources with evidence of shifts in education preferences and practices, for information that resides outside the traditional frames of CCD and IPEDS.

Examples of shifts include changing student/parent choices: Alternative as well as familiar learning environments at all levels (school/home, informal/formal, degree/non-degree/non-traditional options at K-12 and post-secondary levels); Non-degree credentials.

Further transitions include: Teacher mobility and career paths; Changing demographics that alter the geographic distribution of families and students, either permanently or temporarily in response to environmental or other disruption followed by allocation of recovery resources with the COVID pandemic as prime example.

## Π

Guiding principles for undertaking this augmentation of NCES data are: to set a process in place for prioritization, to balance the costs and benefits of potential linkages/file integrations, and to start now with existing resources and opportunities.

#### Define a Process for Prioritization

• Establish guidelines and criteria for requests for data linkage/integration.

These guidelines need to encompass the justification, the resource requirements and the plan (whether tool, reference data set or integrated data) as well as the criteria for evaluation of the request.

The guidelines also need to cover requirements for use of data sets created by other researchers; specifically, both access to files for use by other researchers with appropriate security clearances, and extraction of descriptive or aggregate data files for less restricted access. For both tools and files, full documentation should meet guidelines to be prescribed by NCES (not just instructions to tool or file users).

• **Balance** Issue Importance with Effort/Resources/Time.

Beyond the impact of decisions dependent upon the linked files, a second aspect of importance is the breadth of user community able to access these data. And added to the other side of the balance are the continuing costs, user support, curation and maintenance of files and tools.

With few exceptions, aggregate data and/or small area estimates suffice and reduce negotiation time and data security requirements of micro-data. Notable exceptions are certain longitudinal data collections seeking access to national records of wage, earnings, employment, and student loan data insofar as is permissible by law.

• Attend to Privacy and Data Confidentiality.

Privacy protection requires both operational and technical attention, addressing not only individual person information but also small aggregates or highly cross-referenced data. This is a critical and increasingly complex item as technology continues to advance. Both privacy and confidentiality issues become more complicated as the scope of information is expanded. When

questions can be answered above the level of individual person, linking aggregate external data to NCES data, either micro or aggregate, makes sense.

## Maximize Effectiveness of Existing Research Resources

• **Promote** use of EDGE and other NCES tools.

Highlight EDGE capabilities (with examples) on the website. Develop user resources to support research, for example, to support (automatically) data linkage between different levels of geography standardly used for federal data.

Encourage the inclusion of maps and other graphics (via EDGE and other tools, with acknowledgment) in NCES reports and presentation.

• Improve the Website for Locating Research Resources.

Rethink, then restructure, the website to organize and to advertise the research resources available; these include data files, software tools, research results including publications (from blogs to refereed journals) with good key word search capabilities, as well as research operations information such as project approval requirements for data access. Include an easy to find dashboard of existing research projects, researchers, and the data sets being used (including those linked to NCES data), in order to foster collaboration among researchers.

• Facilitate Creation and Sharing of Link-expanded NCES files.

Partner within the Department of Education to encourage researchers' utilization of linkexpanded NCES files and to leverage Department resources to support access.

Partner with other federal agencies to determine areas of overlapping interest and information; find ways to partner with states where they hold the relevant primary information on high priority common interests.

Provide integrated data files at tiered security levels. Both detailed files for research and files open to broad public access are needed with accessible accompanying maps and graphics.

## III

File integration and tool development require significant technical skills, navigation of a number of hurdles and cooperation with other agencies, none of which should be underestimated.

## Invest the Resources in Expertise and Effort Required.

• **Prefer,** as a general practice, tools for linkage over single purpose data sets.

The flexibility and broad applicability of tools that can extract and create data sets to meet specific demands favor tool development. However, this must be balanced against any increased requirements for their use that limit access by researchers and other users, either in terms of requisite technical expertise or in terms of security clearance for access.

• **Expand** and strengthen geo-based tools for linking data from multiple sources, federal and non-federal.

Particular tools that use geography to index data elements offer access to many different federal (and other) data banks that contain contextual information related to NCES data. Other agencies are also developing tools for data linkage across different geographic units; so partnering in development may offer economy of effort.

• Address the need for increased investment of effort and technical skill for implementing data linkages.

Teaming is essential. For geographically indexed data, both cartographer and statistician are needed; for data indexed by other factors, the required skills are data science (specifically data structure/data linkage) and statistics. A minimum of 2.0 FTE (1.0 for each team member) in addition to current staff/contractors are needed to get started.

- **Develop** and enhance long-term inter-agency relationships around high priority common interests so they are sustainable with the goal of expanding the scope for of shareable information, micro-data and data-linkage and analysis tools.
- Eliminate or work to overcome systemic hurdles.

Technical hurdles include asynchrony in time, scale and mapping unit, non-overlapping or non-adjacent samples and unstructured data. Operational hurdles include data ownership, file permanence, updating and maintenance and management of access at (tiered) security levels.

## APPENDICES

Appendix A: Agenda

- Appendix B: Charge to Panel
- Appendix C: Case Studies
  - Case Study Income Data
  - Case Study Geographic Information

Appendix D: Use Case Scenarios

- Criminal Justice and Safety
- Geographical Information
- Environmental Health
- Teacher Pipeline
- Post COVID Playout

Appendix E: Expert Panel Biosketches

## NATIONAL INSTITUTE OF STATISTICAL SCIENCES

## NCES EXPERT PANEL ON SETTING PRIORITIES FOR FEDERAL DATA ACCESS TO EXPAND THE CONTEXT FOR EDUCATION DATA

## February 17 & 22 & March 2 & 10, 2021

## AGENDA

## Wednesday, February 17

11:00 am – 12:30 pm	Open Session	
11:00 am – 11:15 am	Welcome and Introduction of Panel and NCES staff	
11:15 am – 11:30 am	Review of Charge and Presentation of Work at NCES	
11:30 am – 12:30 pm	Questions from the Panel and Discussion	
12:30 pm – 1:00 pm	Break	
1:00 pm – 3:00 pm	Panel Executive Working Session (closed)	

## Monday, February 22

11:00 am – 3:00 pm	Panel Executive Working Session (closed)	
11:00 am – 12:30 pm	Panel Deliberations (closed)	
12:30 pm – 1:00 pm	Break	
1:00 pm – 3:00 pm	Panel Deliberations (closed)	

## Tuesday, March 2

11:00 am – 3:00 pm	Panel Executive Working Session (closed)	
11:00 am – 12:30 pm	Panel Deliberations (closed)	
12:30 pm – 1:00 pm	Break	
1:00 pm – 3:00 pm	Panel Deliberations (closed)	

## Wednesday, March 10

11:00 am – 1:00 pm	Open Session
11:00 am – 12:00 pm	Review and Presentation of Findings by Panel
12:00 pm – 1:00 pm	Questions from NCES Staff and Discussion
1:00 pm – 1:30 pm	Panel Executive Working Session (closed)

## APPENDIX B: CHARGE TO PANEL

#### Setting Priorities for Federal Data Access to Expand the Context for Education Data

Education policy and education research often depend on a richer context than just education data in NCES data collections. The Evidence Act establishes a <u>presumption of accessibility</u>\* for Federal statistical agencies such as NCES to have access to any Federal data collections where access is not specifically prohibited by law. This panel is charged with setting priorities for enabling integrated perspectives that combine NCES education data with data from *other* federal data collections. Priorities will be viewed from the view of enhancing primary education information and also from the view of external objectives that require education information for analysis or decision-making (e.g., child health/welfare, juvenile justice, labor force/labor market analysis, environmental/hazards data).

\*: Presumption of access actually applies less to statistical data sets which have legal prohibitions than to administrative data sets. In theory, NCES should have presumed access to any administrative data held by any Federal government agency.

### **APPENDIX C: CASE STUDIES**

## Linking Income Data from the Federal Tax System with Education Data<sup>6</sup> Barry W. Johnson

#### INTRODUCTION

National Center for Education Statistics (NCES) survey data provide rich, high quality information on students, teachers, and other educational professionals at a single point in time or over relatively short panel periods, such as 3 to 6 years. In addition, the Department of Education hosts rich administrative data files that can be used to support many important research and evidence-building projects. While these data are very useful by themselves, some important uses require additional socioeconomic data. A few examples of such topics would include measurement of and trends in educational disparities, the effects of education or education policies on long-run outcomes, the incidence of student debt and effects of student debt policies, the interplay between Title IV Federal student aid and education-focused tax policies and Institutional characteristics of for-profit schools. For these topics, some measure of income is needed.

Individual income information provided to the Internal Revenue Service is a possible source of data for studying both students and teachers. These data include earnings; employment including secondary job; migration, including migration across states; and family formation and childbearing. Tax data on businesses and information reported to the IRS by non-profit organizations can provide insights on the finances and operations of academic institutions. Third party data, provided to the IRS from academic institutions on Form 1098-T include post-secondary undergraduate enrollment and tuition/fee payments at the many institutions that are not well covered by current surveys; actual tuition/fees paid and by whom (family, grants); and graduate and mature-age enrollment. These data can be used to give context to how students place in the income and financial aid distribution of their postsecondary institutions and can be more complete, more representative or more accurate than data collected through surveys or even on Department of Education administrative files.

While income data from the tax system offer many advantages, the IRS is prohibited by law from sharing these data except for the specific purposes described in the Internal Revenue Code.<sup>7</sup> These are primarily limited purposes that advance tax administration, including policy development and evaluation at the federal, state, or local level. The IRS Statistics of Income Division is also permitted to release aggregate statistics to the public and to conduct special statistical studies upon request if the outputs do not disclose federal tax information for any specific individual or business. The 2018 Foundations for Evidence-Based Policymaking Act did not include provisions that expanded the use of IRS data. However, as the following sections will describe, there are several possible models for bringing together tax and education data in ways that can advance evidence-building and policy evaluation.

<sup>&</sup>lt;sup>6</sup> Carolyn Hoxby (Stanford University), Sarah Turner (University of Virginia), and John Friedman (Brown University) substantially contributed to the suggestions presented in this case study.

<sup>&</sup>lt;sup>7</sup> <u>Title 26 of the United States Code</u>

## Aggregated or Synthetic Income Data

While individual level microdata is important for many purposes, there are cases where aggregated information or synthetic data added to NCES survey or administrative data would provide valuable context and insights and represent a significant improvement over current research methods, which often require researchers to use statistical methods relying on strong assumptions to, for example, understand education gaps between students from high vs low-income backgrounds. The Department of Education's College Score Card web application provides a useful case study of how aggregate data can be combined with Education data to provide valuable insights.

The College Score Card is an online tool, created by the Department of Education, for consumers to compare the cost and value of higher education institutions in the United States. The data can also be downloaded for research purposes. The Score Card provides data for five areas: education cost, graduation rate, employment information including earnings 10 years after graduation, average amount borrowed, and loan default rate. The earnings data are provided by Statistics of Income (SOI) at the IRS. Data are available for all Integrated Postsecondary Education Data System (IPEDS) institutions that participate in Title IV programs. For each student, the Department provides SOI with an SSN, Institution Code (OPEID), Degree Code, and Program Code (CIP) for gradation cohorts 10 years after graduation. SOI links the SSNs to current year income data and the creates reporting groups by Institution/Degree/Program to compute the following aggregate statistics:

- number with no income information in the tax system,
- number with income above and below the poverty threshold,
- median earnings.

To protect privacy, SOI uses a differential privacy approach to add noise to the data, with the amount of noise based on the number of observations in each reporting group. The resulting count data are quite accurate and the average error for the income data was approximately 2 percent. Importantly, SOI can provide the Department with error measures for each Institution/Degree/Program- level statistic, information that is valuable when using differentially private data for analyses. Once the income data are delivered, SOI destroys the IPEDS data.

This approach is very useful for purposes where income is adding important context to a dataset but where small differences between observations are not critical for the research method. For example, this approach may not work well for research using modern empirical methods such Regression Discontinuity, where small differences in income between records are critical. For many other cases, where *income level* is enough, this approach is promising. Aggregate values could also be used to place survey respondents into context with the full population – for example how do the teachers who respond to a survey fit into the entire population of teachers in the same county or state?

A related approach, that of developing useful synthetic microdata on income that could be linked with NCES survey files for research purposes is also worth exploring. The process could start by directly linking an NCES record to one or several years of tax records filed by a survey respondent. These data would then be used as the basis for imputing a value that is analytically useful but preserves privacy. This would be similar to the approach researchers commonly use when continuous variables are reported in ranges on a

survey. Developing acceptable technical approaches for producing high quality synthetic data could enable many statistical agencies to share restricted use data, such as Title 13 (Census), data and suggests this might be an area for interagency collaboration with leadership from NCES.

An important factor in prioritizing datasets is how difficult logistically it would be to link the NCES data to a tax record, which is the starting point for aggregation or imputation. Surveys for which there are SSN identifiers available for all or most participants would be much easier logistically. Also, surveys that include participants when they are adults (even if they were teenagers when initially surveyed, as occurs with certain longitudinal surveys) are much easier to match based on names, birth dates, and address information. Maiden or otherwise changing surnames are less likely to be an issue when respondents are adults, and participant's address is more likely to match with an IRS information return such as a Form W-2. NCES surveys for which this approach might be appropriate include:

- 1. National Postsecondary Student Aid Study (NPSAS)
- 2. Beginning Postsecondary Students Longitudinal Study (BPS)
- 3. Baccalaureate and Beyond (B&B)
- 4. National Teacher and Principal Survey (NTPS)
- 5. Schools and Staffing Surveys (SASS)

*Cost/barriers:* For the College Scorecard, linking the tax and ED files, computing the required statistics, and processing the data to protect individual privacy costs SOI about \$80,000 per year including overhead, with 25% of that cost devoted to disclosure limitation. This is significantly lower than the cost of producing these data for the first time because much of the programming code can be reused from year-to-year. The Department of Education reimburses SOI for these costs.

#### **INFORMED CONSENT**

Informed consent is an integral part of responsible data practices because it permits providers to retain control over their personal, confidential information.<sup>8</sup> This seems particularly important when using data provided for one purpose, such as tax administration, to support other purposes including research unrelated to tax administration. Under the Internal Revenue Code, taxpayers can authorize the IRS to release tax information, either verbally or in writing, for the type of tax and years specified, to any individual, corporation, firm, organization, or partnership. The process requires taxpayers to file IRS Form 8821. While this provision is most frequently used taxpayers for income verification purposes (for example as part of a mortgage application), it can be used to obtain income data from consenting survey respondents. The University of Michigan Institute for Social Research (ISR) has for many years used this process to obtain earnings histories for consenting respondents to waves of the Health and Retirement Survey (HRS).

In the case of the HRS, SSA processes the IRS Form 8821 because SSA maintains the longitudinal income data on behalf of the IRS. The data provided include information from the Social Security Master Earnings file, Master Beneficiary Record, Supplemental Security Record, and other files. Consenting survey respondents compete an IRS Form 8821 that ISR has prefilled with the requested data items and years. This process can be used to obtain data from already filed tax returns and for a reasonably limited number of

<sup>&</sup>lt;sup>8</sup> The Report of the Secretary's Advisory Committee on Automated Personal Data Systems, Department of Health, Education and Welfare (1973), Records, Computers and the Rights of Citizens Publication No (OS) 73-94 (https://www.justice.gov/opcl/docs/rec-com-rights.pdf) proposed a set of Fair Information Practices Principles that formed the foundation of the Federal Privacy Act of 1974, and remain relevant to current thinking about privacy and confidentiality.

future tax years. Arrangements should be made in partnership with SSA and the IRS before asking respondents to complete the Form 8821 to ensure the consent requests meet all legal requirements.

*Cost/barriers:* Historically, the IRS has required that the Form 8821 include a 'wet' signature to be valid. Respondents participating in online or phone surveys may not be willing to take the additional time and effort to sign and return a paper form, limiting the attractiveness of this approach. However, the Taxpayer First Act of 2019 requires the IRS to develop a process for accepting electronic signatures on this and other IRS forms, which may make this a more viable option in the future.<sup>9</sup> SSA can charge a fee to process these requests based on the number of cases and the size of the resultant data files.

## **BUILDING PARTNERSHIPS TO ACCESS RESTRICTED DATA**

Even in cases where restricted use data, such as tax or Census data cannot be provided to NCES for general purposes, there may be important research or evidence building questions that would satisfy statutory requirements if worked in partnership between NCES and the data owner. The tax system includes several benefits designed to support post-secondary education including the American Opportunity Tax Credit, the Lifelong Learning Credit and the Deduction for Tuition and Fees. Research for which one of the goals is to better understand the impact of these provisions by looking at longitudinal outcomes could benefit both the IRS and NCES.

Carolyn Hoxby and coauthors have demonstrated the potential of this approach through a series of papers that linked tax data with NCES data to provide important insights into both the impact of these tax programs and broader education policy questions.<sup>10, 11, 12</sup>

Each of the papers relied on linking tax data from Form 1040 (including Forms 8863 and 8917), Form 1098-T, and the NCES Integrated Post-Secondary Education Data System (IPEDS). Overall, the findings provided important insights on the efficacy of the tax credits and suggested practical implications for the administration of the tax benefits, including potential form revisions, possible data coordination with other federal agencies, and novel uses of earnings data to target compliance reviews. Thus, each of these papers were demonstrably relevant to tax administration. The data also allowed the authors to develop insights about the educational experience of students as well. For example, Hoxby (2017) provides a detailed profile of students who attend online only for-profit secondary institutions, showing how these students differ in important ways from students enrolled in traditional brick and mortar college programs. Relative to the latter, online students tend to be older, working, and more prone to withdraw from courses without completing them. The analysis also suggests that the return on investment for these online students was low and possibly negative.

Other potential areas for linking IRS and Education data to support research include looking at the interplay between the American Opportunity Tax Credit and Pell Grants, and for past students, the interplay between student loan interest deduction and income-based repayment. This would require linking data from the Free Application for Federal Student Aid (FAFSA) to income tax data. Research looking at the incidence of student debt and the effects of student debt policies, including debt forgiveness has relevance to both tax

<sup>&</sup>lt;sup>9</sup> See <u>https://www.irs.gov/about-irs/electronic-signature-options-will-simplify-third-party-authorizations.</u>

<sup>&</sup>lt;sup>10</sup> Hoxby, Carolyn (2017) "Online Postsecondary Education and the Higher Education Tax Benefits: An Analysis with Implications for Tax Administration," <u>https://www.irs.gov/pub/irs-soi/17rppostsecondaryedtaxbenefits.pdf</u>.

<sup>&</sup>lt;sup>11</sup> Hoxby, Carolyn and George Bullman (2015) "How Tax Credits for Higher Education Affect College-Going," <u>https://www.irs.gov/pub/irs-soi/15rptuitionfeesdeduction.pdf.</u>

<sup>&</sup>lt;sup>12</sup> Bullman, George and Carolyn Hoxby (2014) The Returns to the Federal Tax Credits for Higher Education," <u>https://www.irs.gov/pub/irs-soi/14rpfedtaxcredithighered.pdf</u>.

administration and education policy. This could be achieved by linking income data to administrative data in the National Student Loan Data System (NSLDS) or the National Postsecondary Student Aid Study (NPSAS).<sup>13</sup> In particular, the NSLDS data can make important contributions to both NCES surveys and tax administration by giving context to how students place in the income and financial aid distribution of their postsecondary institution and providing crucial information about how students choose to pay off (defer, default on) their loans, given the numerous choices available to them.

*Cost/Barriers:* SOI has not historically charged fees when a non-IRS research partner participates in a joint research project. In most cases, the non-IRS research partner enters into a no-cost agreement under the Intergovernmental Personnel Mobility Act (IPA) to access tax data.<sup>14</sup> All access to IRS data has traditionally taken place in an IRS facility using IRS equipment. It can take 6-9 months from project approval to complete the required background check, provision equipment, issue an ID badge, and establish the needed computer accounts. If the research partner uses data external to the IRS and those data require special security accommodations, getting those in place can take considerably longer. IPA agreements are for 2 years but can be extended for a total of 4 years.

## FULLY SYNTHETIC DATA

There is a great deal of interest in expanding access to large, administrative files held by the Department of Education because of the ability of researchers to study important subgroups which may not be well represented in NCES surveys. However, for some administrative files, access and use may be legally restricted. As noted above, the NSLDS and FAFSA data are two examples of high value, restricted-use Education administrative datasets that would be very useful to both NCES and other agencies. One approach to increasing the research use of these files would be to create fully synthetic versions that could be used alone or statistically matched to other large datasets to address important research questions. The utility of the synthetic files would be even greater if the actual data could be remotely accessed by researchers through a validation server.

As an example, SOI has engaged with the Tax Policy Center at the Urban Institute (Urban) to create fully synthetic individual income tax databases.<sup>15</sup> The goal is to create files that preserve important statistical characteristics of the 'real' data without containing any identifiable tax return information. To compliment these data, Urban is also developing a procedure that would allow researchers to submit statistical programs that have been tested on the synthetic data to run on the 'true IRS, with results subject to a review that will guarantee that the output satisfies required disclosure avoidance protocols. This 'validation server' approach may open some restricted data to new research uses, because as envisioned, it does not involve granting researchers direct access to sensitive micro data, only to aggregated results protected using an approach based on differential privacy concepts. The Urban project is being guided by a team of experts from within and outside government.

Urban has explored several synthesis techniques including parametric and nonparametric models. A particularly promising nonparametric method, Classification and Regression Trees (CART), sorts observations into relatively homogeneous groups and draws from the empirical distribution of outcomes that occur for each group. There are computational and analytical challenges in implementing this method

<sup>&</sup>lt;sup>13</sup> A suggested way of doing this would be to use the existing 4% sample from NSLDS (the CEAD-STAB) with earnings or Census records an annual basis.

<sup>&</sup>lt;sup>14</sup> See <u>https://www.opm.gov/policy-data-oversight/hiring-information/intergovernment-personnel-act/</u>.

<sup>&</sup>lt;sup>15</sup> For a detailed description of this project, see: <u>https://www.urban.org/research/publication/safely-expanding-research-access-administrative-tax-data-creating-synthetic-public-use-file-and-validation-server/view/full report.</u>

on a large scale, but it appears to be a good option for certain discrete variables. Sequential regressionbased methods are being used to synthesize continuous variables like the sequential regression multiple imputation (SRMI) model that has been used to impute missing values by many researchers.

The resulting synthetic tax products will be useful for many purposes and may even be of interest to NCES for statistical matching with its surveys. Perhaps as important, all the computer code used to create the synthetic tax data and validation server, along with useful documentation, will be made public as part of this project. This will enable other agencies to build on what is achieved through this initiative.

This approach has promise as a potential data "access tier" that could offer a new choice for researchers. NCSES should consider how it might leverage this approach as part of a future comprehensive data dissemination strategy that includes expanded research access to rich Department administrative files that are not currently accessible.

Costs/barriers: This multi-year project is being entirely funded by foundation grants, including a grant from the National Center for Science and Engineering Statistics. It involves several fulltime Urban staff, as well as regular collaboration with SOI staff.

## National Center for Education Statistics Geographic Information

Michael Ratcliffe

## **O**VERVIEW

Demographic, economic and geographic data from several important federal national surveys is available now for schools, school districts, and school neighborhoods by these data files via linkage through geographic data identifiers.

Much of the basic geographic information needed to link datasets inside and outside NCES on common levels of geography or relationships between various levels of geography is already available on NCES's website through the Education Demographic and Geographic Estimates (EDGE) Program and tools, like SAFE Map (Schools Are Far Enough), that the EDGE team has developed.

These data and relationship files, however, are difficult to locate. The first step is to enhance and increase visibility into the range and location of geographic information, particularly the school, school district, and geospatial data available from EDGE. This can be accomplished in part through clear links and pointers on NCES's website, including the main page at <u>https://nces.ed.gov</u>.

More effective highlighting of the existence of these resources will help data users link school district-level data to statistical data available at other levels of geography as well as the legal and statistical geography within which each school is located. These data also provide the base for geospatial and geostatistical methods that facilitate more effective linkage of statistical and other datasets needed to understand the social, demographic, economic, and geographical context affecting students, families, teachers, and schools.

## **NCES RESOURCES**

## Education Demographic and Geographic Estimates (EDGE)

The EDGE Program <u>https://nces.ed.gov/programs/edge</u> offers access to demographic, economic, and geographic data for school districts by drawing on survey and data resources with national coverage. Using the geographic information available through EDGE can maximize linkage through geographic data identifiers and to support data linkage across agencies. In this way, the EDGE Program also serves as an example of inter-agency collaboration as it integrates staff and resources at NCES and the Census Bureau to produce a diverse set of data products.

Data available through the EDGE Program include:

- Demographic data from the 1990, 2000, and 2010 decennial censuses as well as the American Community Survey (ACS).
- Economic data include poverty estimates for the neighborhoods around school locations as well as the comparable wage index for teachers (CWIFT).
- Geographic data, including school district boundaries, school locations, school attendance area boundaries (through the 2015-2016 school year), locale code boundaries, ZIP Code tabulation areas (ZCTAs) as well as variety of geographic area relationship files. Specific geographic data offerings include:

- Archive of district boundaries providing geodata and entity IDs for spatial and tabular connections. NCES creates a single composite layer from the multiple Census Bureau boundary files so that all districts are available in a single file (which is more typical of the way data users work with district data). <u>https://nces.ed.gov/programs/edge/Geographic/DistrictBoundaries</u>
- School geocode files (public, private, postsecondary, and district admin offices) provide options for connecting datasets. These files include point geometry, address geocodes, locale assignments, and a host of additional geocodes that associate the points with various types of legal and statistical geographic areas.

https://nces.ed.gov/programs/edge/Geographic/SchoolLocations

- The EDGE Program's geographic reference file collection provides crosswalks for the complete set of spatial associations between districts and other legal and statistical geographic areas. The geographic reference file collection facilitates identification of the exhaustive set of relationships between districts and other areas for data users who are unable to use as geographic information system (GIS) to do so. It is also important to note that similar files providing relationships between school districts and other types of geographic areas are not available on the Census Bureau's website; as a result, NCES offers the most comprehensive set of geographic relationship files related to school districts. https://nces.ed.gov/programs/edge/Geographic/RelationshipFiles
- The Common Core of Data includes the federal ID and state ID for each school and district. This creates another way of linking local and state school/district information to everything in the EDGE collection.

The EDGE program also provides open access to geographic and socio-economic data through an ArcGISbased platform (available at <u>https://data-nces.opendata.arcgis.com</u>). Functionality at the site includes ability to download data layers, create maps, and use the EDGE open data application programming interfaces (APIs), facilitating integration of data from the EDGE site directly into an application.

EDGE is widely used; more than 125,000 requests for data via EDGE's geographic data web services were received during the week of February 21-27, 2021.

The EDGE Open Data site provides all the NCES geospatial data (and some of the linked ACS district data) as web services to make them easy to find and easy to use in a GIS. The Open Data platform also provides APIs for all geospatial data layers, along with a map viewer that allows users to integrate NCES data with other spatial data layers available in ArcGIS Online (e.g., Housing and Urban Development (HUD) data; Census Bureau data; Federal Emergency Management Agency data, etc.). The EDGE program data layers also are included in Esri's Living Atlas, providing GIS data users another point of access to NCES' geographic information. <a href="https://data-nces.opendata.arcgis.com/">https://data-nces.opendata.arcgis.com/</a>

In addition, NCES school/district geospatial data are included in Data.gov and the Federal Geographic Data Committee's Geoplatform as another way to help data users find, use, and integrate the NCES geodata.

## SAFE MAP SAFE MAP (ed.gov)

The discussion above mentioned EPA's use of NCES's school locations data in its EJSCREEN mapping tool. The EDGE team produced SAFE Map, an interactive mapping tool, to give users the ability to visualize individual school locations and measure the distance to an address. As a tool designed for a specific purpose, its functionality currently is limited to the ability to search on an address, display school locations, and through a customizable radius measure, identify schools within the specified distance from the address of interest.

## EXAMPLE

A fundamentally geographic issue of importance to school districts is the siting of industrial plants that produce pollutants and hazardous waste sites and disposal. Concerns for student exposure have led to laws, regulations, license, permit policies and monitoring requirements for environmental hazards and industrial sites that produce pollutants.

The Environmental Protection Agency's (EPA) environmental justice mapping tool, EJSCREEN, accesses school locations using EDGE's open data API to create the map in Figure 1. In addition to the map, that overlays schools and hazardous sites, this tool links to a report of the data underlying the map.





Using SAFE Map to focus on the near neighborhood of the school, the actual distance from the nearby oil refinery to the school can be calculated as shown in Figure 2.



Figure 2. Distance of Garyville/Mt. Airy Math and Science Magnet School to Oil Refinery Site.

## RECOMMENDATIONS

## First, Enhance and Increase the Visibility of EDGE and the Ease of Access to EDGE and EDGE Tools.

More effective highlighting of the existence of these resources will help data users link school district-level data to statistical data available at other levels of geography as well as the legal and statistical geography within which each school is located. These data also provide the base for geospatial and geostatistical methods that facilitate more effective linkage of statistical and other datasets needed to understand the social, demographic, economic, and geographical context affecting students, families, teachers, and schools.

## Expand SAFE Map Functionality.

Broader capability for integration of additional data would provide NCES data users and stakeholders with integrated platform for linking through the mapping platform to underlying data sets and for visualizing that information. In this way information related to school safety, environmental conditions, and the socioeconomic context where schools are located can be investigated and illustrated. Clarity about the specific questions that need to be answered is the key to effective expansion of data and functionality within SAFE Map for topics of interest that lend themselves to visualization in mapped form.

Expansion of SAFE Map and/or development of similar applications that facilitate mapping NCES and related data as well as download of data and summary reports, will require additional staff resources.

## APPENDIX D: USE CASE SCENARIOS

-			
Domains a	ND CLASS OF QUESTIONS	NCES DATA	External Sources
Domain:	Criminal Justice and Saf	ety	
Questions:	School and Student Safe	ety	
		SSOCS, FRSS, ED <i>Facts,</i> ECLS-K	BJS, CDC, CHDS (Navy), Census, IRS, SSA
			State longitudinal data
Domain:	Environmental Influence	es (Physical/Geographical)	
Questions:	Pollution, Health, and Lo	ong-Term Student Outcome	S
		ELS:2002 & similar	EPA, NCSER
		iongituumai studies	Public school records
Domain: Non-traditional Education and Non-D		on and Non-Degree Credent	ials for Work
Questions:	Economic Value of Industry-Recognized and/or Non-Degree Credentials for Work		Degree Credentials for Work
		NTEWS, IPEDS, NHES, also NPSAS, B&B BPS, HSLS	Census, IRA, SSA, HHS (NDNH)
			Title IV Schools, Industry Credentialing Bodies
Domain:	Education Careers and La	abor Markets	
Questions:	Teacher Pipeline – Prepa	ration and Mobility across	States
		NTPS, IPEDS	IRS, Census, ED:OPE
			State Administrative Data
Domain:	Effects of Disruptions		
<b>Questions:</b> Short- and Long-Term Post-COVID Learning Delay and Losses		d Losses	
		NAEP, PISA	CDC
			States COVID Incidence Data, District level closing Data, Public school records

## DOMAIN: CRIMINAL JUSTICE AND SAFETY

#### Class of Questions: School and Student Safety.

Linking student health and long-term earnings outcomes with school safety issues.

IES/NCES Data: SSOCS, FRSS, EDFacts, ECLS-K, SCS supplement to NCVS, EDGE.

#### **Justification for Priority**

If students do not feel safe or are not safe at school, that could have a negative effect on their ability to learn, and perhaps their probability of landing in the criminal justice system. Other social and health consequences stemming from their exposure to bullying and other behaviors that threaten their safety and self-esteem levels, could be as drastic as loss of life through suicide.

The problem of school violence is prevalent. In 2019, CDC's nationwide <u>Youth Risk Behavior Survey (YRBS)</u> was administered to high school students across the United States. According to the survey results from 13, 677 students: about 1 in 5 high school students reported being bullied on school property, and more than 1 in 12 high school students reported being cyberbullied in the last year; about 9% of high school students had not gone to school at least 1 day during the 30 days before the survey because they felt they would be unsafe at school or on their way to or from school; and about 48% of students in grades 7– 12 experienced some form of sexual violence at school during the 2010–11 school year.

#### According to the CDC's Technical Package on Youth Violence

(https://www.cdc.gov/violenceprevention/pdf/yv-technicalpackage.pdf) "Preschool enrichment with family engagement programs can lower the prevalence of problems in early childhood, including aggression and child abuse and neglect, and have broader and long-term impacts on parent-child interactions and youth's academic achievement, substance use, and perpetration of violence and crime. Examples of effective programs are Child Parent Centers (CPCs) and Early Head Start (EHS). CPCs have been evaluated in multiple, long-term studies. For instance, when followed to age 20, low-income minority children who participated in the CPC preschool program, relative to youth in other early childhood programs, had significantly lower rates of juvenile arrest (16.9% versus 25.1%), violent arrests (9.0% versus 15.3%), and multiple arrests (9.5% versus 12.8%).105 At age 24, relative to youth with fewer years of CPC participation (e.g., preschool only), youth who participated in the program for 4 to 6 years had a 22% reduction in arrests for violence as well as significantly lower rates for violent convictions and multiple incarcerations. Across studies, youth participating in CPCs also experience numerous other benefits relative to comparison groups, including lower rates of substantiated reports of child abuse and neglect, out-of-home placements, grade retention, special education services, depression, and substance use as well as higher rates of high school completion, attendance in four-year colleges, health insurance, and full-time employment in adulthood.

The CDC further finds that:

"The evidence suggests that universal school-based programs can reduce aggressive behavior, including bullying, and other risky behaviors associated with youth violence. Multiple systematic reviews of various universal school-based programs demonstrate beneficial impacts on youth's skills and behaviors, including delinquency, aggression, bullying perpetration and victimization, and bystander skills that lower the likelihood of violence and support victims. For example, the Task Force for Community Preventive Services found a 15% relative reduction in violent behavior among students in pre-kindergarten through high school. Using different outcome measures, the median relative reduction in aggression and violent behavior associated with universal school-based programs varied by grade level, with a 32% reduction for pre-kindergarten and kindergarten students, 18% reduction for elementary students, 7% reduction for middle school students, and 29% reduction for high school students.

#### **Data Linkages**

NCES and BJS produce annual *Indicators of School Crime and Safety*, which is a compilation of data from numerous surveys including NCES conducted *School Survey on Crime and Safety* (SSOCS), *Fast Response Survey System* (FRSS), *EDFacts, the Early Childhood Longitudinal Survey Kindergarten* (ECLS-K) panels. The included data sets also include surveys conducted by other agencies such as the *National Crime Victims Survey* (BJS), *Youth Risk Behavior Survey* (CDC) *National Vital Statistics* (NCHS), *K-12 School Shooting Database* (Center for Homeland Defense and Security, DOD), and the *School Associated Violent Death Surveillance System*.

Consequently, there is value to be added if these key data sets on crime and safety in schools are linked with each other and with geographic and census data to provide better insights into the characteristics of areas most at risk and how various interventions have achieved improvements in those geographic areas. Including engaging parents earlier (such as at the preschool level) and instituting other programs throughout preschool through high school to reduce violence would add further values. The use of different time frames and age categories, for example, complicates this linkage.

A good place to start making improvements may be with the existing administrative records, namely the *School Associated Violent Deaths Surveillance System*, the *National Vital Statistics*, and the *K-12 School Shooting Database*, which are ongoing reporting systems. The *School-Associated Violent Death Surveillance System* (SAVD-SS) — sponsored by the U.S. Departments of Education and Justice as well as the CDC — tracks lethal violence (i.e., homicides, suicides, and legal intervention deaths) on school grounds or on the way to and from school. The *K-12 School Shooting Database* has mapped every shooting incident at a school with information on type of incident, location and whether the shooter was a student. The National Vital Statistics provides the most complete data base on births and deaths in the U.S. Linking these records with Census demographic and geographic data and making it available through EDGE as a geographic layer could add a lot to existing information about school violence and improve the annual Indicators report as well as become a longitudinal file.

Of note, the National Institute of Justice has looked at this issue and stated, "At this time, there is no single data collection that captures the complete picture of the frequency, incidence, and trends in violent crime in U.S. schools. Rather, government agencies and nongovernmental organizations employ numerous data sources and surveys. Some of this information is presented in *Indicators of School Crime and Safety* (the *Indicators* report), which is published regularly by the U.S. Department of Education's National Center for Education Statistics and the U.S. Department of Justice's Bureau of Justice Statistics. The report establishes reliable indicators of the current state of school crime and safety across the nation and is helpful in tracking specific indicators over time; however, it also contains an amalgamation of information on school safety

that is not easily interpreted. This is compounded by the lack of agreed-upon focus and definitions across the sources from which the *Indicators* report draws its data, the irregularity of the data collections, the different populations surveyed (e.g., students versus principals), and differences in how questions are phrased."

To improve on this situation, the report recommends several ways that agencies could coordinate better to improve the data available and also mentions a major new effort it has initiated with several researchers to create an open-source database that includes all publicly known school shootings resulting in at least one firearm death or injury that occurred on school grounds in the United States from January 1, 1990, to December 31, 2016.

NCES Data Survey/Study	External Agency	File(s)	Other Data
SSOCS, FRSS, ED <i>Facts,</i> ECLS-K	BJS	National Crime Victims Survey (NCVS)	State longitudinal data
SCS supplement to NCVS, EDGE		National Vial Statistics System (NVSS) School Associated Violent Death Surveillance System (SAVD-SS, with ED and DOJ) Youth Risk Behavior Surveillance System (YRBSS)	
	CHDS (Navy)	K-12 School Shooting Database	
	Census	American Community Survey (ACS)	
	IRS, SSA	Wage data	

## DOMAIN: ENVIRONMENTAL (PHYSICAL/GEOGRAPHICAL) INFLUENCES

Class of Questions: Pollution, Health, and Long-Term Student Outcomes.

**IES/NCES Data:** NLTS2 (NCSER), ELS:2002, and similar longitudinal studies.

## **Justification for Priority**

A growing body of research suggests that pollution may have pronounced effects on student health and achievement, both in terms of contemporaneous exposure in school (Heissel, Persico, & Simon, 2019; Gilraine, 2019), and in utero exposure (Bharadwaj, Gibson, Zivin, & Neilson, 2017; Sanders, 2012). At the same time, students with disabilities in Special Education may be up to 4 times as likely to have a chronic health condition compared to their peers (Lipscomb et al., 2017; NCHS, 2006). Additionally, the rate of chronic health conditions varies substantially within the different disability categories used in Special Education, with, for example, students with emotional disturbance disorders being 2.7 times as likely as students with a specific learning disorder to have a chronic health condition (Lipscomb et al., 2017). This is concerning for many reasons; for example, for the general student population, a substantial literature finds associations between chronic health conditions and achievement & graduation (Champaloux &Young, 2015; Crump et al, 2013). Yet, very little work has focused on the effects of air pollutants. This proposed linkage would connect EPA data from the Air Quality System (AQS) and the Toxics Release Inventory (TRI) program to previously collected NCSER and NCES data, by exploiting geographic identifiers from both files. In particular, see: <a href="https://aqs.epa.gov/aqsweb/airdata/download\_files.html#AQI">https://aqs.epa.gov/aqsweb/airdata/download\_files.html#AQI</a>

#### Data Linkages

EPA data on pollution is geographic - providing measures such as PM2.5 levels from the AQS and 770 individually listed chemicals and 33 chemical categories collected by the TRI program from different monitoring sites, as well as latitude and longitude. These coordinates can be mapped to NCES public school universe data, which can in turn be mapped to the specific schools, districts, and states that have students followed in relevant NCSER and NCES longitudinal studies. For example, data could be used from students age 13 through 16 in 2000 receiving special education services in grade 7 or above surveyed in the NLTS2 and students in ELS:2002 who are in the same cohort. Data would flow from the research team via their access to publicly available EPA data (AQS and TRI) to NCES, with NCES to perform the linkage of EPA data to the educational data sets, and provide this file to the research team.

NCES Data Survey/Study	External Agency	File(s)	Other Data
ELS:2002 and similar longitudinal studies	EPA NCSER	Air Quality System (AQS) Toxic Release Inventory (TRI) NLTS2 and similar longitudinal studies	Public school universe records to connect EPA geography to state, district, and school locations

## DOMAIN: NON-TRADITIONAL EDUCATION AND NON-DEGREE CREDENTIALS FOR WORK

**Class of Questions:** Economic Value of Industry-Recognized and/or Non-degree Credentials for Work.

Measuring return on investment of post-secondary non-traditional education and of adult training credentials.

IES/NCES Data: NTEWS, IPEDS, NHES, also NPSAS, B&B, BPS, HSLS.

## **Justification for Priority**

In 2009, President Obama's State of the Union Address highlighted the various forms of post-secondary certification and their importance. Concurrently, the deepening recession was leading to an increased research and policy interest in the role of industry-conferred and -recognized credentials in preparing adults for specific, living-wage jobs. Under President Trump, interest persisted, particularly in the area of apprenticeships. Increasingly, jobs are being created in STEM that do not require a 4-year degree, or that require degree recipients to obtain additional, non-degree, credentials. Currently issues of return on investment of various education and training persist, especially with the rising cost of post-secondary education.

As a result, a decade of work within the Federal Statistical System, led by NCES, focused on developing and implementing survey items to measure the prevalence and key characteristics of industry-recognized certifications and licenses, through The Interagency Working Group on Expanded Measures of Enrollment and Attainment (GEMEnA). Questions developed through GEMEnA were included in ELS, NPSAS, BPS, and B&B at NCES, and the CPS and SIPP at Census. <u>https://nces.ed.gov/surveys/gemena/</u>. However, given the importance, particularly for the skilled technical workforce, NCES and its partner the National Center for Science and Engineering Statistics, continued the work, in 2020, launching the National Training, Education, and Workforce Survey (NTEWS) to measure (1) The education, training, and career pathways of skilled technical workers, and (2) the prevalence and interplay of education (postsecondary degrees and certificates), work credentials (certifications and licenses), and work experience programs among American workers.

#### Data Linkages

The full data needed to study the views various credentials, their relevance to employment, and their financial return on investment straddles the questions easily included instruments by NCES (such as NHES which already collects detailed educational information) and those included in instruments by agencies such as Census (such as SIPP, that already collects detailed employment information). Currently both NCES and Census incorporate some GEMEnA questions in various surveys but do not directly share other survey items.

At Census, some GEMEnA questions were included in the Survey of Income and Program Participation. Data on Certifications, Licenses, and Educational Certificates. The nine questions added to the 2012 wave of the 2008 SIPP including the type of certificate/license, the relation to current work, and the reason for acquiring certification (in addition to the standard demographic, occupational, and income questions) <u>https://www.Census.gov/prod/2014pubs/p70-138.pdf</u>. At NCES they were prominently used with the Adult Training and Education Survey (ATES) in the 2016 National Household Education Surveys Program (NHES:2016). The survey was designed to see the type of work credentials and postsecondary educational certificates held by non-institutionalized adults ages 16-65 who are not enrolled in high school, and to gather their perceptions of the usefulness of the credentials. Unfortunately, the response rate was below NCES standards.

<u>https://nces.ed.gov/nhes/ates.asp</u>. This response resistance, after serious efforts to follow up with nonrespondents, underscores the need to compile information from multiple sources for direct information and for imputation using covariates and small-area estimation techniques.

## **Effort and Hurdles**

- Much of the data on credentials is held by professional associations and state licensing bodies administrative data. Credential Engine, for example, has been working towards a model of standardization. <u>https://credentialengine.org/</u>.
- 2. College administrative data on certificates and non-credit courses is vital, but many institutions don't adequately track sub-baccalaureate and non-credit course data. Because schools don't get "credit" in IPEDS for noncredit enrollment, it is often not tracked at all. North Carolina attaches money to noncredit enrollment, and not surprisingly has one of the best sets of state data on noncredit enrollment. Generally, schools capture credit courses leading to an educational certificate. Of course, an issue there is that they sometime "award" a certificate without the person getting it know (e.g., for completion of the basic two-year transfer curriculum). They do that to get "credit" in IPEDS for completion, but it leads to a proliferation of certificates that individuals don't see and really aren't valuable in the labor market. Changes to IPEDS could greatly enhance this data collection and ensure awareness of some certifications.
- Any national source of wage and earnings data, whether HHS's National Directory of New Hires (NDNH), SSA or IRS data, or Census's LEHD data all require legislation or other significant changes to make accessible. For example, see the 2016 OMB White Paper at <u>https://obamawhitehouse.archives.gov/sites/default/files/omb/mgmt-gpra/barriers to using administrative\_data for evidence\_building.pdf</u>.

NCES Data Survey/Study	External Agency	File(s)	Other Data
NTEWS, IPEDS, NHES, also	Census	SIPP, CPS	Title IV schools, especially community colleges
NPSAS, B&B, BPS, HSLS	IRS, SSA, HHS (NDNH)	Wage and earnings	Industry credentialing bodies information

## DOMAIN: EDUCATION CAREERS AND LABOR MARKETS

Class of Questions: Teacher Pipeline – Preparation and Mobility across States.

## IES/NCES Data: NTPS, IPEDS.

## **Justification for Priority**

Studies suggest that teacher labor market frictions have an impact on student achievement (e.g., Kim, Koedel, Ni, and Podgursky, 2017<sup>1</sup>). Many of these frictions arise based on the fact that teacher licensure, tenure, and pension systems are state-specific, creating barriers to cross-state teacher mobility (Goldhaber, Grout, Holden, 2017<sup>2</sup>). As a consequence, an unknown number of people drop out of the teacher labor market when they make a cross-state residential move. The number is unknown because, unfortunately the information required to investigate cross-state teacher mobility, and connections to the aforementioned issues, or even teacher salaries, is quite limited. Investigating many of the cross-state teacher labor market issues would require linkages across state administrative databases (Goldhaber, Grout, Holden, and Brown, 2015<sup>3</sup>).

This lack of information may be particularly problematic for getting a national picture of the supply of new teachers, a common topic of concern (Goldberg, 2021<sup>4</sup>). But gaining a better understanding of this early part of the teacher pipeline could be addressed with changes to IPEDS and/or Title II. In particular, it would be helpful to know more about who is preparing to teach, who drops out of the pipeline at what points, and how desirable is teaching as a profession to individuals with different backgrounds (Goldhaber and Holden, 2020<sup>5</sup>). This would in turn aid in understanding the degree to which the pursuit of a public-school teaching career is associated with teacher preparation program or state licensure requirements (e.g., Angrist and Guryan, 2008<sup>6</sup>; Larsen et al., 2020<sup>7</sup>).

Linkages between NCES datasets and state administrative data to IRS and CPS data would allow us to better track what happens to both prospective teachers and teachers who leave a state dataset. We could answer questions like:

- 1) What are the attributes of those who appear to aspire to teach at one point in their postsecondary careers, and how do those attributes change based on compensation or program or state requirements to be eligible to teach (e.g., teacher salaries, edTPA, licensure tests, etc.)?
- 2) How many individuals who train to teach in one state move to another and why?
- 3) To what extent do *relative* (teacher and non-teacher) salaries explain attrition from the teacher labor market for individuals with different attributes or training?
- 4) What is the impact of state teacher employment eligibility requirements on teacher mobility and staffing challenges?

## **Data Linkages**

Links between IPEDs and state administrative data to IRS and CPS data would allow for tracking of teachers through their careers.

Small changes in the IPEDs and Title II procedures would allow consistent data on those who completed teacher education programs. For example, IPEDs counts teacher candidates and certified teachers pursuing advanced degrees the same; it isn't clear whether alternative IHE programs are included in IPEDS data (it is determined by the individual institutions). IPEDs counts degrees completed while Title II could count it as passing the state licensure (see Goldhaber and Holden, 2020).

NCES Data Survey/Study	External Agency	File(s)	Other Data
NTPS, IPEDS	IRS	Employment Information	State administrative data
	Census	CPS	
	ED:OPE	Title II Data	

- 1 https://www.sciencedirect.com/science/article/abs/pii/S027277571630694X?via%3Dihub
- 2 <u>https://journals.sagepub.com/doi/10.1177/0031721717690367</u>
- 3 <u>https://journals.sagepub.com/doi/10.3102/0013189X15613981</u>
- 4 https://www.nytimes.com/2021/03/27/us/COVID-school-teaching.html
- 5 <u>http://caldercouncil.org/understanding-the-early-teacher-pipeline-what-we-can-and-importantly-cant-learn-from-national-data/#.YDqyeC1h3q1</u>
- 6 <u>https://www.sciencedirect.com/science/article/pii/S0272775707000842</u>
- 7 https://www.nber.org/papers/w28158

## **DOMAIN: EFFECTS OF DISRUPTIONS**

Class of Questions: Short- and Long-Term Learning Delays and Losses.

The academic effects of the COVID-19 crisis, learning options used by different types of students, school reopening decisions, and interventions.

#### IES/NCES Data: NAEP, PISA.

#### **Justification for Priority**

The COVID-19 pandemic is greatly affecting American households. The initial response to the public health crisis almost brought the economy to a halt and the unemployment rate jumped to a historical high. In the spring of 2020, the COVID-19 pandemic closed schools throughout the United States. The resulting shift to remote learning created a significant strain on teachers, students, and families across the country. School districts rapidly had to adapt to facilitate remote learning, with potential dramatically different results for different students. In the fall of 2020, schools began to reopen using in-person, hybrid, and remote learning models. However, as reopening decisions were often made by individual school districts, students in neighboring communities may have had very different cOVID-19 crisis, school district reopening decisions and learning options offered and used by different types of students is crucial for the design and policies and interventions to help minimize the effects of the current crisis on students.

#### **Data Linkages**

NAEP has launched the collection of school reopening information through the NAEP 2021 School Survey, which will survey approximately 3,500 schools each month at grades 4 and 8. Although, this data will be extremely valuable to provide a good picture of school re-opening efforts at the moment, other prior data efforts could help complement this data by providing historical information on school reopenings since the Fall of 2020 as well as context of COVID incidence when these decisions are being made. Geographic data on COVID incidence and public-school learning options are available at the census tract and district level, respectively. Earlier opening data can be found at the district level and has already been collected through the pandemic, for example by MCH Strategic Data (data provided until Summer 2021 at https://www.mchdata.com/COVID19/schoolclosings). County level COVID incidence info can be obtained from the CDC COVID Data Tracker (https://COVID.cdc.gov/COVID-data-tracker) or, for example, from the Johns Hopkins Center for Systems Science and Engineering's COVID-19 data repository (https://github.com/CSSEGISandData/COVID-19). Zip-code and/or school level data is available from individual states.

NCES Data Survey/Study	External Agency	File(s)	Other Data
NAEP, PISA	CDC	COVID Data Tracker	Zip-code and school level data from individual states. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. District level school closing/opening data, such as those compiled at MCH Strategic Data. Public school records to connect external data geography to state, district, and school locations.

## **APPENDIX E: EXPERT PANEL BIOSKETCHES**

#### Mary Bohman, Ph.D.

#### Title: Acting Director and Deputy Director of the Bureau of Economic Analysis

Dr. Bohman works with BEA's executive team to expand the agency's statistical programs and to shape the highest priority research projects in furthering BEA's mission. Dr. Bohman is collaborating with BEA's researchers to bolster the impact and the effectiveness of BEA research. And, she has led the way in crafting a strategy to position BEA to be an employer of choice for the years to come. Before joining BEA, Dr. Bohman spent more than 20 years with the Department of Agriculture in numerous positions, leaving her mark on research involving international trade, farm competitiveness, and data-driven policymaking. At the Agriculture Department, Dr. Bohman was the administrator of USDA's Economic Research Service, where over a seven-year period she led initiatives to strengthen the quality of the agency's research and statistics, increase accountability through the development of metrics, and boost employee engagement through improved communications. She was named a Fellow of the Agricultural & Applied Economics Association, the group's most prestigious honor, in recognition for her outstanding contributions to the agricultural economics profession.

#### Dan Goldhaber, Ph.D.

Title: Director of the Center for Analysis of Longitudinal Data in Education Research (CALDER) at the American Institutes for Research, and Director of the Center for Education Data & Research (CEDR) at the University of Washington

Dan Goldhaber uses administrative data to do research that informs decisions about policy and practice at both CALDER and CEDR. Dan's work focuses on issues of educational productivity and reform at the K-12 level, including the broad array of human capital policies that influence the composition, distribution, and quality of teachers in the workforce, and connections between students' K-12 experiences and postsecondary outcomes. Topics of published work in this area include studies of the statistical properties of value-added measures of teacher effectiveness, the effects of teacher qualifications and quality on student achievement, and the equity of teacher distribution.

Dan's research has been regularly published in leading peer-reviewed economic and education journals such as: American Economic Review, Journal of Human Resources, Journal of Policy and Management, Economics of Education Review, Education Finance and Policy, and Educational Evaluation and Policy Analysis. The findings from these articles have been covered in more widely accessible media outlets such as National Public Radio, the New York Times, the Washington Post, USA Today, and Education Week. Dan previously served as president of the Association for Education Finance and Policy (2016-2017), an elected member of the Alexandria City School Board (1997-2002), and as co-editor of Education Finance and Policy.

#### Frederick M. Hess, Ph.D.

*Title: Resident Scholar and Director of Education Policy Studies at the American Enterprise Institute (AEI)* Dr. Hess' work at AEI focuses on K–12 and higher education issues. He is the author of Education Week's popular blog "Rick Hess Straight Up," is a regular contributor to Forbes and The Hill, and serves as an executive editor of Education Next. Dr. Hess started his career as a high school social studies teacher and has since taught at colleges including Rice, Harvard, Georgetown, and the University of Virginia. His books include "Spinning Wheels," "Letters to a Young Education Reformer," "Cage-Busting Leadership," and "A Search for Common Ground." He holds an MA and a PhD in government and an MEd in teaching and curriculum.

#### Barry Johnson, M.A.

## *Title: Acting Chief Research and Analytics Officer and Director of Statistics of Income Division (SOI) at the Internal Revenue Service (IRS)*

Barry Johnson has been serving as the Acting Chief Research and Analytics Officer at the IRS since May 2019. In this role he leads the Research, Applied Analytics and Statistics Division, overseeing efforts to promote and enhance the application of data and analytic solutions to improve IRS operations and mission effectiveness. Since June 2015, Barry has been Director of the IRS Statistics of Income Division (SOI), one of 13 principle federal statistical agencies. As SOI Director, he oversees production of statistical data on most major tax forms and schedules that are used by the Service to support research and analytics, by Congress and the Treasury Department for budget and policy formation, and by the public to better understand the tax system. He also serves as the Statistical Official for the Treasury Department and is a member of the OMB Interagency Council on Statistical Policy where he participates on several executive committees.

Barry was awarded the rank of Fellow of the American Statistical Association in 2016, elected to the U.S. Conference on Research in Income and Wealth at the National Bureau of Economic Research where he serves on the Executive Committee, and serves on the Board of Directors for the National Tax Association.

#### **Shelly Wilkie Martinez**

## Title: Senior Statistician at the Office of Management and Budget, Statistical and Science Policy, Office of Information and Regulatory Affairs

For the past three years Ms. Martinez has provided government-wide leadership on the implementation of the Foundations for Evidence-Based Act of 2018, the Federal Data Strategy, and the Federal Committee on Statistical Methodology. She returned to OMB in late 2017 after 15 months as Executive Director of the statutorily-created Commission Evidence-Based Policymaking. Ms. Martinez was an early member of OMB's evidence team, and led major data quality improvement efforts in the areas of criminal justice, education, and Federal program evaluation. Ms. Martinez also led early interagency efforts to transform and expand the Census Bureau's Research Data Centers program into a Federal statistical-system-wide shared service. Prior to joining OMB, Ms. Martinez spent 10 years at the U.S. Census Bureau and nearly 8 years on Capitol Hill. She is an elected Fellow of the American Statistical Association.

#### Nancy Potok, Ph.D.

#### Title: CEO of NAPx Consulting

Dr. Nancy Potok served as the Chief Statistician of the United States in the Executive Office of the President until January 2020, during which time she served as a Commissioner on the US Commission on Evidence-Based Policy Making and co-chair of the Federal Data Strategy. Prior to that, she served as Deputy Director and Chief Operating Officer of the U.S. Census Bureau; Deputy Under Secretary for Economic Affairs at the US Department of Commerce; Senior Vice President for Economic, Labor, and Population Studies at NORC at the University of Chicago; and Chief Operating Officer at McManis & Monsalve Associates, a data analytics and organizational transformation firm. She has been an adjunct professor and Senior Fellow at The George Washington University and is a Fellow of the National Academy of Public Administration, nonresident fellow at the Urban Institute, and chairs the Board of Trustees for the Institute of Pure and Applied Mathematics at UCLA. She also serves on the Board of Visitors for the University of Pittsburgh School of Computing and Information and is a contributing editor to the Harvard Data Science Review.

#### Michael Ratcliffe, M.Litt.

#### Title: Senior Geographer in the U.S. Census Bureau's Geography Division

Michael Ratcliffe helps lead activities related to the enterprise-wide Frames Program, which is focused on enhancing linkages between the Census Bureau's foundational geospatial, demographic, job, and business datasets. Prior to taking on this role, he led the Geographic Standards, Criteria, Research, and Quality area of Geography Division, and also has led staff responsible for geographic products and mapping. During his tenure at the Census Bureau, he has worked in both the Geography and Population Divisions, with a primary focus on concepts and definitions of places, urban and rural areas, metropolitan and micropolitan areas, as well as other statistical geographic areas. In addition to his work at the Census Bureau, Mr. Ratcliffe is an adjunct professor at George Washington University, where he teaches population geography. Mr. Ratcliffe holds a Master of Letters degree in geography.

#### Gema Zamarro, Ph.D., CEMFI & UNED

Title: Professor and 21st Century Endowed Chair in Teacher Quality at the Department of Education Reform at the University of Arkansas, Director at Charassein, and Adjunct Senior Economist at the USC Dornsife Center for Economic and Social Research (CESR)

Prior to joining the University of Arkansas, Zamarro was a Senior Economist at CESR (2013-2014), an Economist at the RAND Corporation (2007-2013; Adjunct Economist 2013-2016) and Professor of Econometrics at the Pardee RAND Graduate School of Public Policy (2007-2013), as well as assistant professor in the Department of Econometrics at Tilburg University in the Netherlands and a junior research fellow at NETSPAR (Network for Studies on Pensions, Aging and Retirement) (2005-2007). Dr. Zamarro has performed research on applied econometrics in the areas of education and labor economics. She has completed studies on heterogeneity in returns to education, on the relationship between teacher quality and student performance, on the effect of school closing policies on student outcomes, on the properties of value-added methods for estimating teacher quality, on the effect of dual-language immersion programs, and on the causal effect of retirement on health, among others. Her current research focuses on the measurement and development of character skills, determinants of gender gaps in STEM, the study of teacher labor markets and teacher pensions, and gender and education impacts of the COVID-19 crisis.

#### Panel convened by National Institute of Statistical Sciences

#### Nell Sedransk, Ph.D.

#### Title: Director, National Institute of Statistical Sciences-DC

Dr. Nell Sedransk is the Director of the National Institute of Statistical Sciences. She is an Elected Member of the International Statistical Institute, also Elected Fellow of the American Statistical Association. She is coauthor of three technical books; and her research in both statistical theory and application appears in more than 60 scientific papers in refereed journals. The areas of her technical expertise include: design of complex experiments, Bayesian inference, spatial statistics and topological foundations for statistical theory. She has applied her expertise in statistical design and analysis of complex experiments and observational studies to a wide range of applications from physiology and medicine to engineering and sensors to social science applications in multi-observer scoring to ethical designs for clinical trials.

#### Brian Habing, Ph.D.

## *Title: Associate Director for Education Research at NISS, and Associate Professor of Statistics at University of South Carolina*

Dr. Brian Habing is Associate Director for Education Activities and Research working with the DC Office of the National Institute of Statistical Sciences (NISS) and Associate Professor of Statistics at the University of South Carolina. His research has focused on psychometrics and scale construction, with a particular emphasis on multidimensional item response theory. His research focus includes analysis of education statistics and his interests also extend to statistical education, including work with AP Statistics and the development of new courses at the undergraduate and graduate level.

#### Alexi Albert

#### Title: Research Assistant, National Institute of Statistical Science

Alexi Albert is a Research Assistant at NISS, working with the National Center for Education Statistics (NCES). She holds an MS in Mathematics and Statistics with specialization in Data Science from Georgetown University. Prior to graduate school she taught sixth grade math for two years in Eastern Washington State.

#### Ya Mo, Ph.D.

*Title: Research Fellow, National Institute of Statistical Sciences; Assistant Professor, Boise State University* Dr. Ya Mo is a research fellow at the National Institute of Statistical Sciences and an assistant professor of Curriculum, Instruction, and Foundational Studies at Boise State University. She received a dual major Ph.D. in Measurement and Quantitative Methods and Curriculum, Instruction, and Educational Policy Programs, and an M.S. in Statistics from Michigan State University, as well as an Ed.M in TESOL from Boston University. She researches quantitative methods, psychometric measures, and survey statistics; she also applies quantitative research methods to study substantive topics in education, especially large-scale assessments.