Appendix G

Consortium Coordination (3.1.5)

This appendix includes letters of support from Collaborative members.

AMERICAN UNIVERSITY

WASHINGTON, DC

SCHOOL of EDUCATION

June 14, 2019

To whom it may concern:

American University's School of Education (SOE) is pleased to collaborate with the Urban Institute in its bid to serve as the District of Columbia's Education Research Practice Partnership. As defined in by the District of Columbia Education Research Practice Partnership Establishment and Audit Act of 2018 (D.C. Law 22-268), a Notice of Invitation will be released in the second half of 2019.

The SOE has a rich tradition of excellence in academic research, external funding, teaching, and service across its four areas of study: teacher education, special education, education policy and leadership, and international/global education. It is home to the Institute for Innovation in Education and the newly created Center for Postsecondary Readiness and Success. During academic year 2018–2019, SOE initiated the AU/DCPS Teacher Pipeline Project (Teaching Fellows) which aims to increase the number of highly skilled, diverse teachers in some of the most challenged schools in DCPS. Next year, the Project will include 10 DCPS seniors, many from Wards 7 and 8, who will take education courses as part of a dual enrollment component.

The SOE is also home to influential faculty in the fields of education policy and leadership, special education, global education and teacher preparation/curriculum and instruction. As one of the most research productive units on campus, our work ensures that we build knowledge and enhance its impact through engagement with policymakers and practitioners who put our research to use.

In addition to our research portfolio, we have developed strong programming to align with best practices in the fields of teacher and leader preparation. We have developed a teacher preparation transformation initiative including the integration of high leverage practice skills and learning sciences into our current teacher preparation curriculum. The SOE was selected as a participant in the Learning by Scientific Design Network via Deans for Impact and funded by the Chan Zuckerberg Foundation. I am a member dean of the Deans for Impact, a group dedicated to transforming educator preparation.

All in all, SOE is committed and excited about the prospect of collaborating with the Urban Institute as **a partner in research and practice** to ensure equitable outcomes for all students in the District. Thank you for your consideration and please let me know if you have further questions about our work.

Sincerely,

Cheryl Holcomb-McCoy

Cheryl Holcomb-McCoy, PhD
Dean
School of Education

BELLWETHER
EDUCATION PARTNERS

1426 9th Street, NW
Suite 200
Washington, D.C. 20001
September 5, 2019

To whom it may concern:

Bellwether Education Partners is pleased to collaborate with the Urban Institute in its bid to serve as the District of Columbia's Education Research Practice Partnership. As defined in by the District of Columbia Education Research Practice Partnership Establishment and Audit Act of 2018 (D.C. Law 22-268), a Notice of Invitation will be released in the second half of 2019.

Bellwether Education Partners is a national nonprofit focused on dramatically changing education and life outcomes for underserved children. We do this by helping education organizations accelerate their impact and by working to improve policy and practice, through a combination of policy analysis and research, strategic advising, implementation support, and educational program evaluation. While our organization's focus and work are national in scope, we have extensive experience in the District of Columbia: our team has studied education issues and policies in the District (most recently in our Eight Cities analysis of city-based education transformation efforts), provided evaluation support and strategic advising to District-based education organizations (including the Office of the State Superintendent of Education, CASA DC, and numerous charter schools), and translated insights and lessons from the District's experience to inform efforts in other states and nationally.

We are excited to partner with the Urban Institute. As a research partner in this project, we will bring deep expertise in public policy, early childhood education, and evaluation methods. Our practical experience supporting District-based education organizations to improve their educational offerings and reach also gives us a deep familiarity with the on-the-ground operating realities facing schools, families, communities, and children in the District. And our national perspective will enable us to help the partnership place what is happening here in national context and identify and translate lessons from the District that can inform efforts elsewhere.

Thank you for your consideration.

Sincerely,

Sara Mead

Sara Mead
Partner, Bellwether Education Partners

BROOKINGS

1775 Massachusetts Avenue, NW
Washington, DC 20036
telephone 202.797.6000
fax 202.797.6004
web brookings.edu

Brown Center on Education Policy

June 24, 2019

To Whom It May Concern:

The Brown Center on Education Policy at the Brookings Institution is pleased to collaborate with the Urban Institute in its bid to lead the District of Columbia's Education Research Practice Partnership. We understand that a Notice of Invitation will be released in the second half of 2019, as defined by the District of Columbia Education Research Practice Partnership Establishment and Audit Act of 2018 (D.C. Law 22-268).

Founded nearly a century ago, the Brookings Institution is a nonprofit public policy organization based in Washington, DC. Our mission is to conduct high-quality, independent research and, based on that research, to provide innovative, practical recommendations. The Brown Center on Education Policy brings rigorous empirical analysis to bear on education policy in the United States. Brown Center researchers examine an assortment of education policy issues, including issues related to accountability, school choice, and educational equity.

As a Research Partner, we would collaborate on the design and implementation of research studies. Brown Center researchers have considerable experience in designing and conducting empirical studies using student-level data from state departments of education and large urban districts. We would appreciate the opportunity to apply our subject matter and methodological expertise to generate research that supports students in our hometown of Washington, DC.

Thank you for your consideration.

Sincerely,

Jon Valant
Fellow
Brown Center on Education Policy, Brookings Institution

D.C. POLICY CENTER

September 6, 2019

To whom it may concern:

The D.C. Policy Center is pleased to collaborate with the Urban Institute in its bid to serve as the District of Columbia's Education Research Practice Partnership. As defined in by the District of Columbia Education Research Practice Partnership Establishment and Audit Act of 2018 (D.C. Law 22-268), a Notice of Invitation will be released in the second half of 2019.

The D.C. Policy Center is a 501(c)(3) nonprofit think tank advancing policies for a strong and vibrant economy in the District of Columbia. We provide objective, high-quality data analyses to support wide-ranging and productive policy debate in the District. The D.C. Policy Center launched its Education Policy Initiative in 2017 with dedicated funding and staff for education research on the District's public schools, exploring the connections between the city's changing demographics and its public school students. The Education Policy Initiative has developed a data-rich landscape of traditional public and public charter schools in the District of Columbia exploring the determinants of in- and out-of-boundary participation, potential growth paths for public school enrollments, and diversity across District's public and public charter schools. The Education Policy Initiative's research is well-respected by a wide variety of stakeholders, as evidenced by features in local media, views of its reports, and invitations to present at local and national conferences and for testimony at public hearings for the D.C. Council.

The D.C. Policy Center will be engaged in both in research supporting the RPP and act as an intermediary between schools and researchers.

Thank you for your consideration.

Sincerely,

Yesim Sayin Taylor

Executive Director, D.C. Policy Center

EmpowerK12

October 3, 2019

To whom it may concern:

EmpowerK12 is pleased to collaborate with the Urban Institute in its bid to serve as the lead agency for the District of Columbia's Education Research Practice Partnership. As defined in by the District of Columbia Education Research Practice Partnership Establishment and Audit Act of 2018 (D.C. Law 22-268), a Notice of Invitation will be released in 2019.

EmpowerK12 has extensive expertise in school data across all the District's educational stakeholders, including DME, OSSE, DCPS, and Charter LEAs. We know who is collecting what data and with what fidelity. Our organization's vision is that if DC schools dedicate themselves to continuously improving using the best data and research available, the District of Columbia can become the first urban jurisdiction to close the national achievement gap for our students most at risk of academic failure by the end of the next decade.

While a significant portion of the educational data required for a successful RPP will come directly from OSSE via access to SLED and SEDS, much of the on-the-ground data needed to impact daily improvement decisions of teachers and school leaders exists at the LEA and school level. EmpowerK12 creates standardized data warehouses for all our school partners, customized based on their student information system and other data sources. From those data warehouses, we create data dashboards for schools and utilize the data to run robust predictive analytics algorithms designed to provide additional information in the continuous improvement process. Lastly, we support the facilitation of effective data meetings at the school level to ensure accurate interpretation of the data and robust conversation about next steps.

An effective Research Practice Partnership is a pivotal component needed to bend the learning curve of all DC students to new heights. We anticipate supporting Urban Institute and the broader RPP community as a data and practice liaison between the RPP and schools, including facilitating the collection of high quality data directly from schools and supporting schools to continuously improve utilizing research produced by the RPP.

Sincerely,

Josh Boots
Executive Director

empowerk12.org | 202.715.3534 | 800 Maine Avenue SW, Suite 200 Washington, DC 20024

GEORGETOWN UNIVERSITY
Office of Government Relations
and Community Engagement

January 31, 2020

To Whom It May Concern:

Georgetown University is pleased to collaborate with the Urban Institute in its bid to serve as the lead partner and administrative home of the District of Columbia's Education Research Practice Partnership. As required by the District of Columbia Education Research Practice Partnership Establishment and Audit Act of 2018 (D.C. Law 22-268), a Notice of Invitation was released in the second half of 2019.

Georgetown is a leading international research university, and the nation's oldest Catholic and Jesuit university. An anchor institution in the District of Columbia since 1789, Georgetown features many highly ranked academic programs at its four locations in the District of Columbia.

Cutting-edge interdisciplinary research collaboration with the Urban Institute and other members of the Research Practice Partnership will engage Georgetown faculty at many of the university's top-ranked schools and centers, including:

- Georgetown University Medical Center, an internationally recognized academic health and science center with a four-part mission of research, teaching, service and patient care (provided through a partnership with MedStar Health.)
- The McCourt School of Public Policy, top-ranked public policy school which features a diverse community of problem-solvers, committed to moving bold ideas to action. The McCourt School is also home to the Massive Data Institute, which uses data science to seek solutions to societal-scale problems using novel and traditional large-scale data sources.
- Georgetown University Law Center, where theory and practice meet, Georgetown Law features world-class faculty, celebrated theorists and leading practitioners who are actively involved outside of the classroom in writing, interpreting and enforcing the law.
- Georgetown University Center for Child and Human Development, established over 50 years ago and dedicated to improving the quality of life for all children and youth and their families, especially those with special health care needs, behavioral health challenges or disabilities.

Over the past several months, we have worked closely with the Urban Institute to develop a partnership model for the research collaborative which we believe will maximize our ability to help strengthen the District's education system and the academic and life success of its students. The governance structure we have developed together builds on the strengths of each member of the proposed partnership and will balance the individual autonomy of each partner with a desire to make collective impact.

If we can answer questions about the partnership or provide additional information about our work, please do not hesitate to contact me. Thank you for your consideration.

Sincerely yours,

Christopher Murphy
Vice President for Government Relations and Community Engagement

June 25, 2019

To whom it may concern:

The George Washington University Trachtenberg School of Public Policy and Public Administration is pleased to collaborate with the Urban Institute in its bid to serve as the District of Columbia's Education Research Practice Partnership. As defined by the District of Columbia Education Research Practice Partnership Establishment and Audit Act of 2018 (D.C. Law 22-268), a Notice of Invitation will be released in the second half of 2019.

The School has many faculty and students with extensive expertise and reach in the DC metro area as well as an affiliation with the Center for Washington Area Studies (CWAS). CWAS is a university-wide center that conducts research on the neighborhoods and communities of the Greater Washington Area. The Center produces the *State of the Capital Region* report, which provides detailed information on housing growth and affordability in the area. CWAS also convenes events on major policy issues in the region, such as growth in residential segregation and the use of tax incentives to entice business investment.

The Trachtenberg School and CWAS will provide support in the form of faculty expertise and research assistance where needed. Key faculty members include Dr. Dylan Conger, Professor in the School and affiliate of CWAS. Dr. Conger focuses on explaining disparities in achievement between social groups and identifying and improving education policies that reduce those disparities. Dr. Conger's recent research concerns early postsecondary opportunities (such as Advanced Placement) and policies and practices concerning immigrant and English language learner students. Dr. Kathryn Newcomer, Professor and Trachtenberg School Director is a national expert in program evaluation and also available to assist on evaluative projects. The Trachtenberg School and CWAS can also quickly respond to needs by drawing upon the assistance and expertise of the many graduate students receiving their training in public policy, data science, economics, education, and public administration. Finally, the Trachtenberg School and CWAS can assist with organizing and hosting major policy events.

Sincerely,

Dylan Conger

Dylan Conger
Professor, Trachtenberg School of Public Policy and Public Administration
Research Affiliate, Center for Washington Area Studies
George Washington University
805 21st Street NW, Room 601H
Washington, DC 20052
dconger@gwu.edu
202-994-1456

HOWARD UNIVERSITY

School of Education
Office of the Dean

May 23, 2019

To whom it may concern:

The Howard University School of Education is pleased to collaborate with the Urban Institute in its bid to serve as the District of Columbia's Education Research Practice Partnership. As defined in by the District of Columbia Education Research Practice Partnership Establishment and Audit Act of 2018 (D.C. Law 22-268), a Notice of Invitation will be released in the second half of 2019.

The Howard University School of Education, designated in 1971, emerged from Howard University's Normal Department, established in 1867. Our mission is to influence social policies, empower communities, and promote social justice for Black and underserved populations locally, nationally, and globally. We have been an engaged partner in the District of Columbia through projects such as the Partnership for Early Engagement in Computer Science, family engagement support at the Ron Brown College Preparatory High School, and the Center for Drug Abuse Research. We are accredited by the Council for the Accreditation of Educator Preparation and ranked as a top 100 Best Education School by the U.S. News and World Report Rankings.

Our faculty and staff are comprised of a mix of strong researchers, who conduct large- and small-scale research, as well as highly-skilled practitioners across three Departments: Curriculum and Instruction, Educational Leadership and Policy Studies, and Human Development and Psychoeducational Studies. Our faculty have developed national reputations in areas such as education policy, Black education, Latino leadership, school safety, higher education policy, and trauma-informed schooling. Moreover, we have scholar-practitioners who have served as licensed school psychologists, counselors, teachers, principals, superintendents, and executive-level leaders in education. Thus, we believe that we can offer a range of research and practitioner support to the District of Columbia Education Research Practice Partnership. Thank you for your consideration.

Sincerely,

Dawn Williams, Ph.D.
Dean

2441 4th Street, NW
Washington, DC 20059

Telephone 202 806 7340
Facsimile 202 806 7018
www.howard.edu

Sheena McConnell
Senior Vice President

1100 1st Street, NE, 12th Floor
Washington, DC 20002-4221
Telephone (202) 484-9220
Fax (202) 863-1763
www.mathematica-mpr.com
(202) 484-4518

April 11, 2019

Matt Chingos and Megan Gallagher
Urban Institute
500 L'Enfant Plaza SW
Washington, DC 20024

Dear Dr. Chingos and Dr. Gallagher:

Mathematica is pleased to partner with the Urban Institute on its proposal for a Research-Practice Partnership in the District of Columbia (DC). Our ongoing work in DC education has engaged stakeholders and brought researchers and practitioners together to make evidence-based decisions and improve educational outcomes. We have engaged with decision makers in DC to solve real problems and conducted research to respond to important needs. For more than 10 years working in DC education, we have built up a reservoir of trust and respect for the quality and objectivity of our work. The depth and breadth of our collective corporate experience suit our proposed roles, including helping to craft the research agenda, creating a data warehouse, cleaning and maintaining data, creating public-use files, conducting research activities, and disseminating findings.

Over the past decade, Mathematica has worked closely with the District of Columbia Public Schools (DCPS), the DC Office of the State Superintendent of Education (OSSE), the DC Education Consortium for Research and Evaluation (EdCORE), foundations, and the U.S. Department of Education (ED), including the Regional Educational Laboratory Mid-Atlantic (REL Mid-Atlantic), to support DCPS and DC charter schools in making data-driven decisions, engaging stakeholders, and disseminating research findings.

Across 14 recent projects in DCPS and DC charter schools, Mathematica staff have collected and analyzed data on student achievement, enrollment and attendance, and background characteristics; classroom rosters and course-taking; educator performance measures; school lottery applications; disciplinary incidents; and health outcomes. In Exhibit 1 in an attachment, we summarize this work and our deep experience using these data sources. Mathematica's work in DC includes the following features:

- For DCPS and OSSE, we developed assessments, provided technical assistance, and evaluated programs. For example, Mathematica staff have worked with DCPS since 2009 to develop and improve IMPACT, its teacher assessment system, including designing and implementing its teacher and school value-added models and providing technical assistance to effectively interpret and use the results.
- In leading the REL Mid-Atlantic, Mathematica has partnered with state and local agencies to identify high-leverage issues and increase the partners' capacities to access, conduct, interpret, and apply research to those issues. This work is driven by the needs of educational agencies, rather than the interests of researchers, who may not always be primarily focused on solving real world problems. In response to challenges identified by DCPS and OSSE, Mathematica has undertaken analyses and evaluations for the REL on a variety of topics. For example, we supported DCPS to validate measures of socioemotional learning, study home visiting, educate students with chronic health conditions, and require Advanced Placement courses. Currently, we are supporting OSSE to develop and improve school report cards and high school growth models.
- As an engaged partner to DC-EdCORE, Mathematica supported efforts by the DC auditor to evaluate the impact of DC school reforms that began in 2007. As part of this work, Mathematica staff examined trends in teacher retention and effectiveness over time.
- Foundations have supported some of our work on topics of great importance within DC and of broad interest to education researchers and decision makers. This work includes efforts to understand patterns in school choice in DC and the impact of principal dismissals and replacements that occurred after 2007.
- In addition to extended projects, Mathematica staff have engaged with DC stakeholders in several other ways to build relationships and share their expertise. Our staff have presented as experts at DC State Board of Education meetings, DC Council hearings, and have led workshops at the DC Data Summit and throughout the year in FOCUS workshops targeted to data managers in DC charter schools.

From our history of working closely with DC stakeholders, Mathematica offers a team of researchers, statisticians, analysts, programmers, and communications professionals with experience directly related to our proposed role in the research–practice partnership. Many of these staff members work in our DC office. In an attachment, we list the qualifications of several Mathematica staff who have worked closely with DCPS or OSSE; have experience working with data from DCPS or OSSE; or have other relevant experience such as conducting education research, supporting educational decision making, and disseminating research.

Mathematica's experience in DC has given us insight on the direction of new research that can meet the needs of stakeholders and educators. For example, a better understanding of how families make decisions about where to attend school in DC can help schools plan for the future, and support DCPS and OSSE to provide parents with accurate and useful information. Also, our detailed knowledge of administrative data in DC situates us to effectively support efforts to make these data more open and accessible.

We look forward to working with you on this important project to engage stakeholders, identify promising areas for useful research, and disseminate research findings in ways that support understanding and use. If you have any questions, please email rfpcenter@mathematica-mpr.com or, if you need to speak to someone directly, call Pamela Tapscott, Mathematica's vice president of contract operations, at (202) 484-3294.

Sincerely,

S. McConnell

cc: Elias Walsh, Dallas Dotter

ATTACHMENT

Exhibit 1. Relevant project experience

			Disciplinary incidents, social-emotional competencies, or students' health outcomes	Student achievement	Students' school enrollment or attendance	Student background characteristics	Classroom rosters or course-taking	Educator performance measures	My School DC Common Lottery applications
DCPS and OSSE									
Value-Added Assessment System for DC Schools and Teachers (2009–2015)	DCPS & charters	DC CAS	✓	✓	✓	✓			
My School DC Audit (2014–2015)	DCPS & charters		✓	✓			✓		
DC EdCORE									
Impact of 2007 DC School Reforms (2013–2014)	DCPS	DC CAS	✓	✓		✓			
Foundations									
The Impact of Replacing Principals on Student Achievement in DC Public Schools (2013–2014)	DCPS	DC CAS; SAT-9	✓	✓					
Market Signals: A Deep-Dive Analysis of Parental School Choice in Washington, DC (2013–2015)	DCPS & charters		✓	✓		✓	✓		
OSSE School Enrollment Demand Simulator (2017–)	DCPS & charters	PARCC	✓	✓			✓		
Evaluation of DC Education Governance Reforms (2018–)	DCPS & charters	NAEP	✓	✓					
DC Family Engagement Program Evaluation (2014–2016)	DCPS & charters		✓	✓					
Regional Educational Laboratory									
Developing and Validating Socio-Emotional Learning Measures from DCPS Student Survey Data (2018–2019)	DCPS	PARCC	✓						✓
Study of OSSE School Report Cards (2019–)	DCPS & charters	DC CAS	✓			✓	✓		
DCPS Impact Evaluation of Structured Relationship-Building Home Visits (2018–)	DCPS	ANet; DC CAS; PARCC	✓	✓	✓				✓
Supporting Students with Health Conditions in District of Columbia Public Schools (2018–)	DCPS	PARCC	✓	✓					✓
The Impact of Advanced Placement Policies in DCPS (2019–)	DCPS	AP; DC CAS; PARCC	✓	✓	✓				
Technical Support for School Leader IMPACT Evaluation	DCPS	PARCC				✓			
U.S. Department of Education									
Evaluation of Technology Interventions for KIPP-DC (2016–2017)	Charters	MAP	✓	✓	✓				

ANet = Achievement Network; AP = Advanced Placement; DC CAS = District of Columbia Comprehensive Assessment System; DCPS = District of Columbia Public Schools; MAP = Measures of Academic Progress; NAEP = National Assessment of Educational Progress; PARCC = Partnership for Assessment of Readiness for College and Careers; SAT-9 = Stanford Achievement Test, 9th edition.

Qualifications of Mathematica staff

Elias Walsh (Ph.D., Economics, University of Michigan) is a senior researcher at Mathematica who has worked closely with DCPS and OSSE on multiple research projects and has broad expertise in education research and dissemination. An expert in measuring educator effectiveness, Dr. Walsh designed many features of the value-added models DCPS used in its IMPACT evaluation system, and he supported DCPS staff to use and understand the results. Leading Mathematica's partnership in DC-EdCORE, he directed research studies on trends in retaining effective teachers and the effects of replacing school principals in DCPS. Dr. Walsh has also worked closely with policymakers in several other states and districts to develop and use value-added models in systems that evaluate educators. He currently directs a What Works Clearinghouse (WWC) project that reviews effectiveness research for students in early childhood

to high school and disseminates the findings. Dr. Walsh has co-authored many Mathematica reports, working papers and briefs, and has published his work in *Education Finance and Policy, Economics of Education Review,* the *Journal of Research on Educational Effectiveness,* and *Statistics and Public Policy*. Before joining Mathematica in 2011, Dr. Walsh was a Teach For America math teacher in the Chicago Public Schools.

Dallas Dotter (Ph.D., Economics, University of California, San Diego), a researcher at Mathematica, has extensive experience studying education policies in DC. His work includes studies of the impacts of the 2007 DC school reforms on student achievement, teacher effectiveness and retention, and the effect of principal replacements during this time on student outcomes. Dr. Dotter has worked closely with My School DC to study parents' preferences for schools and evaluate the lottery's success in achieving its intended matching techniques. He currently leads a project that combines detailed school data with DC school application data to provide local education agencies with policy simulation tools that will inform them of likely movements of students across schools in response to future school-planning decisions.

Duncan Chaplin (Ph.D., Economics, University of Wisconsin–Madison) is a senior researcher at Mathematica with extensive experience evaluating education interventions, including six programs implemented in DC. He is currently senior advisor for a study examining DC education reforms since 2007. As a principal investigator, he led the design and analysis for evaluations of the DC 2004 summer school program and the DC 21st-Century Community Learning Centers. He also led the development of value-added models to measure educator effectiveness in DCPS. Dr. Chaplin has prepared public-use files, including one for the DC mayor's office on services for DC youth and another for the Millennium Challenge Corporation that combined data from many administrative and survey sources. Dr. Chaplin has co-authored half a dozen briefs and published widely in academic and nonacademic journals including the *Journal of Policy Analysis and Management,* the *Journal of Research on Educational Effectiveness,* and *Economics of Education Review*.

Mary Grider (M.B.A., Yale School of Management) is a senior systems analyst and director at Mathematica with expertise leading system development and processing of administrative records from school districts for numerous education projects. In her work with DCPS and OSSE, she led a team of programmers who prepared student and teacher data for analysis and developed systems and programming code to implement value-added models and report results. She carefully documented data requirements and business rules describing decisions, including the eligibility of teachers, schools, and students for the value-added model. Ms. Grider has also prepared restricted-use data files and documentation for many studies for the Institute for Education Sciences, including for studies of the Teacher Incentive Fund and Access to Effective Teaching for Disadvantaged Students.

Emma Ernst (M.A., Mathematics, Boston University), formerly Emma Kopa, is a senior systems analyst at Mathematica who is deeply familiar with data from DCPS and OSSE. As deputy project director of Mathematica's work developing value-added results for DCPS and OSSE, Ms. Ernst worked extensively with administrative data, prepared data requests, documented data requirements, developed business rules, and supported DCPS and OSSE staff to use the results. Ms. Ernst also led a team of programmers to study the Access to Effective

Teaching for Disadvantaged Students, preparing and analyzing administrative data and producing value-added results for 26 districts. She has created restricted-use files for several studies, including Studies on Streamlining the Certification and Verification Processes of Children from Low-Income Households.

Allison McKie (Ph.D., Economics, Massachusetts Institute of Technology) is a senior researcher with a proven record of collaborating closely with school districts, foundations, and other stakeholders to conduct high quality evaluations and provide evaluation technical assistance. Dr. McKie established a strong working relationship with DCPS when designing and administering a teacher survey and as principal investigator for a REL Mid-Atlantic study on the impact of structured teacher home visiting on student and teacher outcomes in DCPS. She is an expert in preparing, evaluating, and supporting educators, having served as co-principal investigator and project director for the Evaluation of the Teacher Advancement Program in Chicago; a technical assistance provider to grantees for the National Evaluation of the Teacher Incentive Fund sponsored by ED; and a leader of the Teacher Training, Evaluation, and Compensation topic area for the WWC.

Ignacio Martinez (Ph.D., Economics, University of Virginia) is a researcher at Mathematica with experience working directly with decision makers to help them use data to inform their choices. In developing prototypes for predictive analytics dashboards, Dr. Martinez interviewed and collaborated with school leaders in DCPS and DC charter schools, and led an interactive workshop for the DC Data Summit. He has helped school districts develop and implement customized plans to evaluate the effectiveness of educational technologies, designed and implemented rapid-cycle evaluations of educational technologies, and applied his expertise in using Bayesian statistics for decision making to developing the Rapid Cycle Evaluation Coach for the Office of Educational Technology at ED.

Alyson Burnett (Ph.D., Measurement, Statistics, and Evaluation, University of Maryland), is a researcher at Mathematica with 10 years of experience conducting education research, including research for DCPS and DC charter schools. She is a co-principal investigator for the REL's impact evaluation of AP course policies in DCPS. Before this task, she co-led site visits to KIPP (formerly known as the Knowledge Is Power Program) DC schools as part of an evaluation of leadership practices within KIPP. Dr. Burnett has led implementation analyses on research–practitioner partnerships with Atlanta Public Schools and with Say Yes to Education in Syracuse, New York. She is a co-author of practitioner-friendly reports and briefs, including the recent guide funded by ED, Presenting School Choice Information to Parents: An Evidence-Based Guide.

Paul Burkander (Ph.D., Economics, from Michigan State University) is a researcher at Mathematica with experience working on large national education research projects, including federally funded evaluations of the Teacher Incentive Fund and Youth CareerConnect. Dr. Burkander contributed to an evaluation of DCPS's New Heights, a program that delivered school-based services to pregnant and parenting teens; he currently contributes to the analysis of DCPS family engagement programs and is co-principal investigator on an evaluation of DCPS AP policies. Dr. Burkander has experience communicating and disseminating findings from

these analyses to broad audiences, including school district and program staff and federal project officers.

Menbere Shiferaw (Ph.D., Public Policy, New York University) is a researcher who specializes in using applied methods to study education policies and practices. She has more than eight years of experience analyzing K–12 and postsecondary education data. Dr. Shiferaw currently works with OSSE to test and validate alternative approaches for calculating high school growth models and with DCPS to improve how IMPACT evaluates school leaders. She has worked with the New York City Department of Education to strengthen formal and informal education collaborations. Currently, as technical assistant liaison to the department, she helps grantees with varying backgrounds build capacity to understand the applied research process.

Natalya Verbitsky-Savitz (Ph.D., Statistics, University of Michigan), a senior research statistician at Mathematica, has more than 10 years of experience in evaluating programs and policies in K–12 education. She has played a key role in several evaluations of charter schools and charter management organizations, including the National Evaluation of Charter Management Organizations, and is currently involved in evaluating three KIPP school leadership training programs. More broadly, she has contributed to evaluations of literacy programs and policies intended to improve school climate and decrease negative behavioral outcomes, such as out-of-school suspensions and expulsions and teen pregnancy. She currently assists ED to improve its policies and practices on sharing restricted-use data with researchers.

Jeffrey Terziev (M.P.P., Georgetown University), a Mathematica research analyst, has experience working closely with DCPS, conducting education research and communicating the results. Mr. Terziev worked closely with DCPS staff to design a study on home visiting programs under the REL Mid-Atlantic. Before joining Mathematica, Mr. Terziev completed two internships with DCPS, with the assessments and the teacher recruitment and selection teams. His efforts to communicate research findings to a broad audience include a fact sheet explaining three common measures of student growth for teacher evaluation and an infographic on the benefits of improved teacher diversity.

John L. Czajka (Ph.D., Sociology, University of Michigan) is a senior fellow at Mathematica with 40 years of policy research experience. A recognized expert in statistical uses of administrative records, Dr. Czajka has contributed to the design of cross-sectional and panel databases of tax returns and led the application of statistical disclosure limitation to producing public-use files. Dr. Czajka's research in education has included serving as a reviewer for the WWC and REL projects conducted for ED.

Joanne Pfleiderer (B.A., English, State University College of New York at Buffalo), the director of communications at Mathematica, is a versatile communicator with expertise in education. She is skilled at translating complex research effectively for diverse groups and developing websites and interactive online content. An award-winning writer, she is the author of guides on effective dissemination and Section 508 compliance for the RELs. She provides technical support for dissemination projects funded by ED and leads dissemination and engagement activities for the REL Mid-Atlantic.

Lindsay Ochoa (M.P.A., Lyndon B. Johnson School of Public Affairs, The University of Texas at Austin), a research analyst at Mathematica, brings expertise in conducting research projects for education clients and communicating with school and district leaders. Ms. Ochoa previously worked for an education research firm for which she completed research projects for 73 K–12 education clients, including state and local education agencies. She worked with the client and content director to identify the primary research needs, develop and propose methodologies for addressing the research questions, and carry out the research agenda. For Mathematica and other organizations, Ms. Ochoa has co-authored numerous public-facing briefs designed for a variety of audiences, including researchers, practitioners, and policymakers.

Charles Tilley (M.P.P., Public Policy, University of Virginia) is a senior programmer at Mathematica with experience in quantitative analysis, data wrangling and validation, public use files, and large-scale data collections. He was a lead programmer for preparing and documenting the restricted-use files for the National Longitudinal Transition Study 2012, an analysis of disabled and nondisabled youth as they transition from high school, sponsored by ED. He has developed generalizable diagnostic tools to assess data integrity and authored centralized coding resources to automate data preparation for multiple major data collections, including the Impact Evaluation of Support for Principals sponsored by ED.

Kathryn Cronquist (M.A., Latin American Studies, School of Foreign Service, Georgetown University) is a senior programmer with project experience in education, nutrition, and human services. Proficient in Stata, SAS, and SQL, she uses her programming expertise to clean and analyze data on various projects. She also serves as a project task lead and co-authored the 2017 report on the characteristics of households and participants receiving benefits from the Supplemental Nutrition Assistance Program. Ms. Cronquist formerly worked as a contracting data analyst at OSSE.

Urban Institute
500 L'Enfant Plaza, SW
Washington, DC 20024

To Whom It May Concern:

I am pleased to submit a letter of support on behalf of Raise DC for the DC Education Research Practice Partnership. I serve as the Interim Executive Director and Senior Director of Postsecondary Initiatives at Raise DC.

Raise DC is an independent, cross-sector partnership that advances educational equity through data-driven strategies in five citywide goals areas in Washington, DC, from birth through age 24. We convene more than 250 partners at all levels of education, business, government, and philanthropy in all of the District's eight wards. Raise DC uses citywide data in all of our workstreams to both examine the intersection of factors such as race and place and to uncover how interventions have led to change over time for our young people. With timely and actionable data, our collective partnership can ensure our efforts are on track and eventually lead to more equitable and meaningful outcomes.

Raise DC lends its support to and advocates for organizations and entities that reflect our shared values on data and its responsible use. These values include:

- A demonstrated commitment to equity and solving for systemic issues that disproportionately affect DC's most vulnerable residents.
- A dedication to publicly available and accessible data that encourages DC's education community to act on outcomes. This means that data is:
    - Inclusive of diverse stakeholders, including parents, families, caretakers, and direct client-serving organizations (including nonprofits and schools);
    - Accessible to all interested organizations – a "pay to play" structure is not employed;
    - Secure and confidential at identifiable levels; and
    - Released at a predictable and ongoing cycle/frequency (at least annually) and, to the extent possible, supportive of continuous improvement efforts.
- Data/research priorities that reflect the values, interests, and needs of diverse stakeholders who are involved throughout the formation process, including families, youth, and children whose voices are represented from the onset of projects that access the repository.
- Data and efforts that are complementary to the existing data work of our other education stakeholders, particularly the Office of the State Superintendent of Education, DC Public Schools, public charter schools, local education agencies, and community-based organizations.
- Data and research that point toward shared citywide goals and measures, to the extent possible – including kindergarten readiness, high school graduation, postsecondary enrollment and completion, reconnection to school and work for those who have dropped out, and career preparedness.

A data consortium that reflects the values outlined above can support our partners throughout the District in considering and constructing strategies that provide equitable opportunities for young people along their educational journeys, with an ultimate goal of readiness for sustainable careers.

Raise DC is eager to partner with organizations committed to data availability and supporting the vision of success for all DC children and youth.

Respectfully,

Tiffini Andorful, PhD
Interim Executive Director / Senior Director of Postsecondary Initiatives
Raise DC

Trinity

Office of the President
Trinity Washington University
125 Michigan Avenue NE, Washington, DC 20017
v 202-884-9050 f 202-884-9056
president@trinitydc.edu www.trinitydc.edu

November 11, 2019

To whom it may concern,

Trinity Washington University is pleased to collaborate with the Urban Institute in its bid to serve as the District of Columbia's Education Research Practice Partnership. As defined by the District of Columbia Education Research Practice Partnership Establishment and Audit Act of 2018 (D.C. Law 22-268), a Notice of Invitation will be released in the second half of 2019.

Trinity Washington University is a comprehensive university offering a broad range of degree programs to a diverse population of students, a majority of whom are DC residents. Trinity's longstanding programs in Education have provided significant preparation for thousands of DC teachers and school administrators. The majority of Trinity undergraduates are graduates of DC Public or Charter Schools, and other local school systems. Trinity's commitment and contribution to the education landscape of the city is also clear from its numerous partnerships. As one of the founding partners of THEARC, Trinity is the first and only private university offering college-level classes east of the Anacostia river. The largest enrolled program at THEARC is the AA in Early Childhood Education, which is critical for ensuring that the early childhood workforce in DC meets the minimum education requirements implemented by OSSE in 2018. Recently, Trinity has expanded its partnership with DCPS to launch the Early College Academy at Coolidge High School. Since the summer of 2018, Trinity has been pleased to collaborate with OSSE to deliver a Summer Bridge Program to college-bound DCPS students. The NASA DC Space Grant Consortium recently funded a partnership between Trinity and McKinley Technology High that provides undergraduate research experience to high school students.

Trinity is uniquely positioned to serve as both a Practice Partner and a Research Partner to the Urban Institute. As an Educational Provider Program in DC with close relationships among our graduates and their colleagues in local schools, we also bring a knowledge of the needs and experiences of DC teachers. In addition, our Teacher Education faculty bring to the partners significant subject matter expertise in Early Childhood, Special Education, and Literacy. As a comprehensive university with a faculty who are active researchers can also contribute methodological expertise for research design, implementation, and analysis.

Thank you for your consideration.

Sincerely,

Patricia McGuire
President

UNIVERSITY OF THE DISTRICT OF COLUMBIA
1851

January 30, 2020

Mathew Chingos

Vice President, Center on Education Data & Policy

Urban Institute

500 L'Enfant Plaza, SW

Washington, D.C. 20024

Subject: Letter of Support for the Proposal by the District of Columbia Education Research Collaborative in support of the District of Columbia Education Research Practice Partnership Establishment and Audit Act of 2018

M. Chingos,

The University of the District of Columbia (UDC) enthusiastically supports the Urban Institute's formation of the *District of Columbia Education Research Collaborative*, and its leadership of the Collaborative's proposal to the District of Columbia (DC) for secure access to data from DC Public Schools. The diversity of partners in the Collaborative brings enormous depth and expertise to analyze the data repository of the District of Columbia's Public Schools. This is extremely important for the District of Columbia, as our country faces increases global competition, and we must increase the pool of domestic talent, especially in the areas of science, technology, engineering, and mathematics (STEM). As there is a race for leadership in artificial intelligence, advanced manufacturing, digital healthcare, and quantum computing, UDC must do its part as the city's only public, urban, land-grant, emerging research university, and with our Collaborative partners to ensure our residents are not underrepresented in these novel STEM arenas. We proudly support this proposal initiative with the Urban Institute, and our Collaborative that we hope will provide research from the available data leading to solutions that will create an eminent cadre of the next generation of innovators and leaders from the District of Columbia!

Sincerely,

Victor R. McCrary, PhD

Vice President for Research & Graduate Programs

Member, National Science Board

Office Vice President for Research & Graduate Programs
Suite 2327
42500 Connecticut Avenue, N.W.
Washington, D.C. 20008

UNIVERSITY OF MARYLAND

3119 Benjamin Building
College Park, Maryland 20742-1121
301.405.2334 TEL 301.314.9890 FAX
www.education.umd.edu

COLLEGE OF EDUCATION
OFFICE OF THE DEAN

May 28, 2019

To whom it may concern:

The University of Maryland's College of Education is pleased to collaborate with the Urban Institute in its bid to serve as the District of Columbia's Education Research Practice Partnership. As defined in by the District of Columbia Education Research Practice Partnership Establishment and Audit Act of 2018 (D.C. Law 22-268), a Notice of Invitation will be released in the second half of 2019.

The College of Education's mission is to enhance the lives of individuals, families, schools and communities through our research, teaching, and engagement. We have a long history of faculty undertaking research with DC schools on improvements in curriculum and professional development and seek to build on that history. Our 2022 strategic plan, as one of its central pillars, seeks to foster our strategic partnerships and formalize long-term collaborative institutional partnerships that focus on educator development, research, school/program improvement, and policy.

If the Urban Institute is awarded a contract to serve as the District of Columbia's Education Research Practice Partnership, we will be glad to discuss how we can engage with the Urban Institute as a research partner and/or a practice partner. In terms of research, our faculty have expertise in evaluation methods, research design, data use and analysis, and psychometrics, as well as content-area expertise in pedagogy, school finance, human development, counseling, and school leadership. In terms of practice partnerships, we have a range of clinical experience in working with families to foster development of children and in the provision of professional development to support teachers, counselors, and other school professionals.

Thank you for your consideration.

Sincerely,

Jennifer King Rice
Dean

UNIVERSITY of VIRGINIA | CURRY SCHOOL of EDUCATION and HUMAN DEVELOPMENT

May 8, 2019

To whom it may concern:

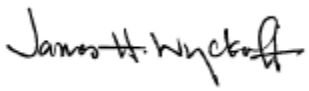
I am pleased to collaborate with the Urban Institute in its bid to serve as the District of Columbia's Education Research Practice Partnership. As defined in by the District of Columbia Education Research Practice Partnership Establishment and Audit Act of 2018 (D.C. Law 22-268), a Notice of Invitation will be released in the second half of 2019.

Tom Dee and I have worked with DCPS over the last eight years on a variety of projects to produce evidence on DCPS policy reforms. That work has been published in a variety of peer reviewed journals and disseminated in the media, including the New York Times, NPR, the Washington Post, etc. I have a strong commitment to producing rigorous research that informs policy decisions.

In the past I have spent a meaningful portion of my time on this work and it has been a source of research for 8 of our PhD students, who have produced research collaboratively and independently. I envision that model continuing.

Thank you for your consideration.

Sincerely,



James H. Wyckoff
Curry Memorial Professor of Education and Public Policy
Director, EdPolicyWorks

STANFORD
UNIVERSITY

April 25, 2019

Dear Colleagues:

I am pleased to collaborate with the Urban Institute in its bid to serve as the District of Columbia's Education Research Practice Partnership. As defined in by the District of Columbia Education Research Practice Partnership Establishment and Audit Act of 2018 (D.C. Law 22-268), a Notice of Invitation will be released in the second half of 2019.

I am a scholar of education policy and have worked with increasing frequency in partnership with school districts and state education agencies, including DC Public Schools. As a Research Partner, I intend to provide both subject matter and methodological expertise relevant to the shared work of the Partnership.

Thank you for your consideration.

Sincerely,

Thomas S. Dee
Barnett Family Professor
Faculty Director, John W. Gardner Center for Youth and Their Communities

520 Galvez Mall, CERAS Building, 5th Floor, Stanford CA 94305-3001
E-mail: tdee@stanford.edu Voice: 650.723.6847 Fax: 650.723.9931

Appendix H

Expertise in Data Security and Management (3.1.6)

This appendix includes descriptions of Collaborative members' data security and management plans.

URBAN INSTITUTE · ELEVATE · THE · DEBATE

Urban Institute Procedures for Protecting Confidential Data

The Urban Institute has developed explicit procedures regarding the protection of confidential data. This can be superseded by specific requirements imposed by funding agencies regarding confidential data.

Overview

The term "confidential data" encompasses any information designated confidential by external agencies or parties with whom Urban has a data use agreement, and information designated as sensitive material by the Urban Institute's Institutional Review Board. All managers of projects using confidential data must complete a data security plan, which includes specification of a data security officer for the project. Only those with a documented "need to know" are allowed access to confidential data. All employees who may access confidential data are required to take IRB training.

Project managers are encouraged to limit access to confidential files as much as possible. Access restriction can be achieved by limiting the use of confidential variables. For example, if a file is considered confidential because it contains identifying names and addresses, those variables may be removed from the file and replaced with pseudo identifiers. The sanitized file can then be used and shared without risk of violating confidentiality. Access restriction can also be achieved by limiting staff members' access to particular computer accounts or an entire set of files.

Although technological tools and procedures can assist in the protection of confidential data, the Urban Institute requires each person who uses confidential data to adhere to the applicable data security plan and ensure that their work habits are secure. All staff members, consultants, and subcontractor staff using or handling confidential data must sign confidentiality pledges asserting that they will adhere to the guidelines for confidential data use and nondisclosure. Project managers reinforce the importance of these pledges, clearly explain the specifics of the data security plan, and monitor their teams' security practices.

Before Receiving Confidential Data

A project manager must submit a Data Security Plan to the IT Security Officer in advance of receiving any confidential data. Arrangements are then made for appropriate transmission of these data:

- File transfer over secure electronic connections. This means the source system must be a trusted and recognized source for the data, and the means of transfer must be secure, such as an encrypted internet connection to the Urban Institute. Urban maintains a secure ftp server where external parties can exchange data with Urban through encrypted connections.

- Delivery by secure, trackable means, such as FedEx, UPS, or registered US mail.
- Hand delivery by a cleared individual.

Urban Institute staff are prohibited from using inappropriate delivery mechanisms, such as unencrypted file transfer over the internet or unencrypted email.

Storing and Protecting Confidential Data
The Urban Institute maintains confidential disks separate from nonconfidential disks on all its system servers. Confidential data are stored on the confidential disks and not backed up in Urban's normal system backups. Staff are instructed not to copy these data to a nonconfidential disk. If confidential data are stored on an individual PC, these data are encrypted or stored on removable storage media that is secured in a locked cabinet when not in use. Urban's standard encryption software is PGP, but specific requirements from project sponsors will be honored.

When a staff member uses an account or a computer with access to confidential data, he or she does not leave the session unattended, logs out at the end of the session, and locks ups any storage media that hold confidential data.

File protections and access controls are established to ensure that confidential data are not accessible to anyone who is not explicitly authorized to use them. More specifically, across operating systems, access control lists grant access only to project members and system administrators.

All storage media (e.g., CDs, internal and external hard drives, flash drives) that hold confidential data are explicitly labeled confidential. Project managers maintain a log for each piece of confidential storage media recording the following:
- Receipt of item from external source
- Creation of item at the Urban Institute
- Destruction of item
- Transfer of item to someone else's responsibility (even within Urban)

If a project requires off-site storage or archiving of confidential data (for purposes of disaster recovery), Urban ensures that the off-site facility is authorized to hold and protect confidential data.

Confidential storage media and printouts can be removed from the Urban Institute only when they are hand delivered to a person authorized to receive them.

Each center at the Urban Institute is required to conduct annual training for its staff regarding the proper handling of confidential data and to conduct a semi-annual review of confidential logs, confirming that all confidential media can be accounted for.

Disposal or Scrubbing of Confidential Storage Media
The acceptable methods for the disposal or "scrubbing" of confidential storage media include returning the media to the source, physical destruction, or erasure using a "secure erasure" product. Confidential printouts are disposed of by shredding. Hard drives are degaussed; CDs and DVDs are shredded.

Infrastructure to Support Data Security
Urban maintains a private LAN on premise and a virtual private cloud (VPC) presence on AWS (Amazon Web Services). Urban strives to maintain data security on all systems. On premise, an industry-standard firewall is maintained, which evaluates and monitors all attempted connections from the internet to internal servers and to our private network. On our AWS VPC, we also maintain a firewall and store files in S3, where files can be encrypted at rest as needed. Data transmissions from our Urban Institute private net travel via an AWS Direct Connect dedicated line and are encrypted during transmission as required.

For security logging and monitoring, Urban has a 24/7 security monitoring service, which proactively allows us to detect, notify, analyze, and report on server conditions. On servers and user workstations, antivirus software is kept up to date, and we employ best practices in our procedures for securing servers, desktops, and laptops. Confidential data that must be stored locally are encrypted using PGP, which provides FIPS 140-2-level protection.

Data Use by Project Teams
All projects' use of confidential human subject data must be approved by the Urban Institute's Institutional Review Board, adhering to the IRB data security guidelines, and data security plans for any kind of confidential data are reviewed by IT Security Officer.

If a data use agreement requires it, all staff using data under the agreement will sign a confidentiality pledge before access to the data will be granted. The IT Security Officer maintains a list of all staff having access to data under a data use agreement and a contact to which to report any incident that occurs under the agreement. Staff who have confidential data access are required to have PGP full disk encryption with FIPS 140-2-level protection. As required under agreements, data are scrubbed of any personal information such as IDs, names, addresses, and birth dates and will be aggregated at a level so that individuals cannot be identified by one or more variables.

URBAN INSTITUTE · ELEVATE · THE · DEBATE

Urban Institute Institutional Capabilities: Information Technology Services

August 2019

Urban's technology and data science (tech and data) department provides data processing and information management. Its expertise spans data extraction and coding, the development of large database-management systems, sophisticated microsimulation models, website construction, and maintenance. We also have Alfresco as our content management system.

Collectively, Urban's professional programmers and analysts are proficient in the C#, C++, Visual Basic, XML, and FORTRAN programming languages; the Oracle, MySQL, and SQL-server database systems; and such statistical software as SAS, Stata, R, and SPSS. Tech and data staff also create interactive web interfaces using Drupal, Cold Fusion, PHP, Java, and JavaScript, which can connect to web-accessible databases in SQL Server, mySQL, and Oracle.

Tech and data staff are currently involved in developing or maintaining numerous applications. Microsimulation applications include the TRIM III client-server model, the dynamic simulation model DYNASIM, and the Tax Policy Center Microsimulation Model. Many internet websites are maintained, as well as an enterprise-wide intranet. Staff members have extensive experience with social science databases, including NCES, CPS, SIPP, Census, Federal Justice Criminal Processing, and Urban's own NSAF and ANF state databases.

In addition to the professional programmer analyst staff, most Urban research assistants operate standard statistical software such as R, SAS, SPSS, and Stata. The tech and data team helps maintain an enterprise-wide SAS users group with training and developmental workshops. Furthermore, many researchers have experience with spreadsheet and database packages such as Microsoft Excel and Access, and with graphic packages such as PowerPoint. Urban maintains commercial survey development software (Qualtrics) and has also created numerous web-based custom surveys, case management and data collection systems.

A Hewlett-Packard DL580 Windows 2008 R2 Server is our heavy-duty SAS server on premise, and we can spin up a large virtual server in AWS for faster processing.

Many network servers reside on a HP SimpliVity/VMware cluster of HP ProLiant DL380 G10 servers. Each server has 90 terabytes of compressible storage and 384 gigabytes of RAM. The network currently supports more than 600 IBM-compatible PCs. All infrastructure systems communicate to each PC via an Ethernet LAN or Aruba wireless LAN. The LAN includes a pair of

core switch which is a HP 5412zl G3 switch operating at 10/100/1000 Mbs, also HP 3910 switches connected via 10Gb fiber backbone on each floor operating at the same speed.

VMware virtualization is heavily utilized. The clusters provide a high level of manageability, flexibility, and redundancy. A HP P7400 iSCSI SAN with a raw capacity of about 147 TB and a Hewlett-Packard P4000 iSCSI SAN with raw capacity of 147 TB provides disk space for our virtual environment as well as for are statistical servers SAS1, STATA2, and STATA3. The SANs consist of multiple nodes each housing a RAID array. Each node is redundant. The SAN is easily expanded by adding additional nodes or drives.

The network provides file, print, and software services. The printers on the network include multiple color multifunction printers on each floor supporting PCL, POSTSCRIPT, and ASCII. Secure printing is available that delays printing until the user logs in at the device. The LAN has disk space for project teams to share files and for users to back up PC fixed disks. The LAN also serves as a conduit for email, using MS Outlook/Exchange/Office 365. All staff may send and receive messages and files (ASCII and binary including formatted text) to individuals inside and outside Urban. The tech and data team also provides web-based secure file transfer (SFTP) for external user collaboration secured with AES-256 encryption.

The Urban Institute runs three internet connections, a high-speed 1 Gbps fiber-optic connection from Cogent Communications, another with Comcast and a Direct Connect to AWS. Using the Border Gateway Protocol (BGP) on our Cisco ASR1001 routers, we have automatic failover between the two internet connections. This not only benefits the Urban Institute staff members, who often need to use the internet in their work, but it also benefits those who visit Urban's web servers or who send email to Urban. All Urban websites are maintained on either Linux or Windows servers, and Urban supports Microsoft IIS, Apache, Tomcat, and containerized web services.

The Urban Institute strives to preserve data integrity and security. A Checkpoint firewall cluster monitors and evaluates all attempted connections from the internet to our public web servers and our private network. A managed security service monitors the network via sensors, agents, and log collection to provide 24/7 security. Up-to-date Symantec and Microsoft Defender antivirus software runs on our desktop PCs and our servers. We also implement other best practices for securing our servers and our desktop PCs. Locally stored confidential data are encrypted using PGP AES-128 encryption, which provides FIPS 140-2 validated protection. All projects' use of confidential human subject data must be approved by the Urban Institute's Institutional Review Board, adhering to the IRB's data security guidelines.

All Urban staff members are supplied with IBM-compatible PCs. The typical PC configuration is a 3.50GHz Dell Precision PC with 16/32 GB RAM, 500Gb hard drive, CD/DVD RW, and 19-inch monitor. Standard software includes Windows 7, Office 2010 Professional (Word, Excel, ACCESS, Power Point), and Microsoft Outlook 2010. Urban also supports more than 80 compatible laptop computers. Remote access to the LAN is available through a Citrix SSL VPN, multifactor authentication is required for access to confidential data.

American University
School of Education

Expertise in Data Security

American University's Office of Research (OR) works with Principal Investigators (PI) to ensure that data are managed in appropriate and acceptable ways that ensure the highest level of security and support for the research function. Working with PIs, the OR first evaluates the requirements imposed by the restricted use dataset, determines if current capabilities and resources can meet them, works with the Office of the Information Technology (OIT) coordinates the provisioning of technology and process to satisfy the requirements and the Office of the Chief Information Security Officer to obtain technology risk advice. The OR assures that any compliance agreements concerning the use of restricted use datasets are signed by a specific representative of the university with signatory authority, the Vice Provost for Research, and works to implement a plan to assure adherence to all data compliance agreements.

American University agrees to store data on systems that meet the following requirements:

- Anti-malware software is installed, running, and updated regularly;
- Requires separate credentials for each user; and
- The hard drive is encrypted using Microsoft Bitlocker, Apple FileVault or similar whole disk encryption software
- Every AU-owned computer is set up to meet the previous security guidelines.

In addition, American University, along with over 750 academic institutions and research organizations, is a member of the *Inter-university Consortium for Political and Social Research (ICPSR)*, which provides leadership and training in data access, curation, and methods of analysis for the social science research community.

Expertise in Data Security

To avoid confidentiality breaches, Bellwether uses secure cloud software (e.g. Qualtrics, Salesforce, and Dropbox Business, which allows encryption) to transfer and store personally identifiable information and other sensitive data. Unique identifiers are only obtained when necessary. Data are separated from identifiers by coding and maintaining separate files, and identifiers are deleted once the unique identifier has served its purpose (e.g. has been used to merge dataset). Two-step verification is used for all data (e.g. passwords on individual documents and folders and the storage site), and data are purged after three years. Specific projects adhere to any additional data security requirements of the agency providing the data. For example, we have a data-sharing agreement with the Department of Education in Louisiana that requires us to access data via a File Transfer Protocol (FTP) only once we are physically in the state, and then upload the data via the FTP and remove any data files from our computers prior to leaving the state.

Bellwether has several processes in place to ensure that any data collected are handled securely. Data collected on portable electronic devices (for example, cell phones, tablets, thumb drives, or laptops) are transferred as soon as possible to secure cloud storage. Recordings that are subsequently transcribed are destroyed after confirming the transcription's accuracy and completeness. Similarly, original paper forms are destroyed (such as by shredding the documents) after data are transferred to a secure electronic format.

EXPERTISE IN DATA SECURITY

Personally identifiable information will be secured at the Brown Center, which has researchers experienced in protecting and working with these types of data. The Brown Center has the following data management policies in place to ensure data confidentiality: personally identifiable or other sensitive data are stored on a non-networked desktop computer in a secure office; the desktop computer is password-protected and the hard drive is stored in a locked safe when not in use; and the original data are not duplicated (only one backup-copy will be stored, in the same manner as the original), accessed only by project personnel and authorized research staff (with signed Affidavits of Nondisclosure as required to become authorized), and used only in the authorized location.

We prevent disclosure of potentially sensitive information through the following procedures: only authorized research staff work with the data; all results and reports are reviewed by at least one scholar with relevant expertise prior to publication; and drafts of all publications are provided to the data provider prior to public release.

Of course, we will also comply with all relevant laws, rules, and agreements in how we handle any data shared with us.

Expertise in Data Security

The D.C. Policy Center uses standard industry practices to provide both hardware and software level encryption. The D.C. Policy Center uses up-to-date operating systems with hard drive level encryption provided through the individual hardware providers. Software encryption is enabled when necessary for the data being handled. The software level encryption is password protected 256-bit key based AES by default. The Center also has internal classification levels that provide a strict framework of access and control of sensitive information, including PII.

When data is stored in house, the data is saved with the software and hardware encryption enabled and behind a controlled access office and controlled access building. If need be, the confidential data can be transmitted using standard encrypted data transfer techniques and stored off site in an encrypted cloud-based data warehouse. The Center follows all federal and local requirements around sensitive data by ensuring all controls are in place and operational. Raw data is not transmitted to unauthorized users for any purpose and authorizes new users through an internal assessment of need and risk. The D.C. Policy Center will not publish or distribute confidential or sensitive data and will only publish data and derivatives of sensitive data after an internal data sanitization process has been completed and evaluated.

EmpowerK12 School Data Security Practices

EmpowerK12 has developed data security expertise by successfully analyzing and keeping secure the data from over 95 schools in the District of Columbia over the last 5 years. We utilize a defense-in-depth approach to security where we restrict access to servers on deny-by-default basis, automatically patch security vulnerabilities, transfer data using encrypted means, monitor the physical security of our on-premises data, and follow a comprehensive protocol for handling students' personally identifiable information. We view the largest threat to data security as human error and, accordingly, follow and frequently revisit our internal controls, training, and norms. EmpowerK12's security elements are summarized in the chart below and described in further detail on subsequent pages.

Summary Chart

Area	**Type**	**Physical**	**Identity & Access**	**Perimeter**	**Network**	**Compute**	**Application**	**Data**
SharePoint/ Teams	SaaS		Two-factor authentication for admins, group-based management					
Virtual Machines	IaaS		Access by admins only	Virtual network, VPN, Azure data security deny by default	Security groups deny by default	Patches, Anti-malware, Antivirus		Encryption
SQL Server	PaaS		Access by admins only	Azure SQL Server firewall deny by default				Encryption
Power BI	SaaS		Row-level security, Audits of Users and Access				Design practices to limit inadvertent exposure to PII and inability to easily copy and share PII	
Local Machines	On-prem	Policy to report stolen or compromised computers and devices, auto-locking devices via password and facial recognition	Single user only, employee-signed data protection agreements	Microsoft Azure monitoring of access and alert system for unusual access		Patches, malware, antivirus		Follow standard backup and retention policies, Encryption

Physical

Physical resources, such as laptops and mobile devices need to be protected from theft and accidental access. They should be kept under the control of EmpowerK12 staff or locked in a secure location. When unattended, devices should be locked and protected via password or biometrics. Devices should be set to auto-lock after five minutes. If a device is lost, stolen, or compromised in any way, the responsible staff member is required to notify EmpowerK12's managing director or executive director.

Identity & Access

While the practices above are intended to secure physical resources, it is possible that a device may become compromised. Because of that, it is critical that sources of student data are protected by requiring authentication and authorization.

Limited Access

Our primary method of keeping data secure is limiting access only to the small number of individuals within EmpowerK12 or data admins at partner organizations. Unlike IT environments with too many users to manage manually, we limit access to virtual machines to only EmpowerK12 staff, and we carefully monitor and adjust access as needed. SQL server access is only granted to a data admin at partner organizations, and again, access is tightly controlled.

Two-Factor Authentication

Because many of our resources are kept in Microsoft services (e.g. Azure, Office 365), we utilize two-factor authentication for Microsoft Office and Azure admins.

Group Policies and Row-Level Security

EmpowerK12 utilizes Microsoft Teams' embedded group-based security and access policies to manage the exchange of information with school partners. When possible, we sync permissions and user accounts with partner databases to reduce management burden and to ensure user permissions are kept up to date. In our Microsoft Power BI dashboards, we take advantage of Power BI's row-level security features to further limit access to student information to only those who need it for an educational purpose.

Perimeter

We maintain a virtual network in Azure and because those virtual machines are used only for ETL purposes, they are tightly locked down with a deny by default filtering policy using Azure's Network Security Groups. We access that vNet via VPN to create a secure tunnel to our local machines. Azure SQL server is another resource that needs protection from potential attacks, and our server is protected by Azure's built-in firewall where we deny-by-default and whitelist only the IP addresses of EmpowerK12 administrators.

Network

Within our virtual networks, machines are not able to communicate with one another unless they are specifically whitelisted. Therefore, if one machine is compromised, the other machines can remain secure.

Compute

In case an attack does get through our previous layers of defense, it is critical that each machine has defense software to detect and stop threats. We set machines, both local and virtual, to auto-update with security patches. We also utilize industry-standard antivirus and antimalware software.

Application

Aside from malicious attacks, student PII may be exposed through inadvertent mistakes, during both the ETL (extract, transform, and load) process and the Power BI app creation process. Staff are trained on common sources of error that may expose data, and we employ code review and Q&A processes to ensure that sensitive data are not shared.

Data

The last layer of defense is encrypting the data we need to protect. We encrypt our SQL Server databases, virtual machine hard drives, and all employee local machine drives. It is also important to follow practices that prevent the inadvertent sharing of PII. File storage practices and e-mail policy described in our corporate Backup and Retention Policy are designed to prevent accidental sharing. All staff members sign privacy agreements to ensure they understand what is considered acceptable data usage and sharing of private information.

Expertise in Data Security

The Massive Data Institute (MDI) at Georgetown's McCourt School of Public Policy focuses on the secure and responsible use of data to answer public policy questions. MDI works with researchers in government, academia, and industry to solves societal-scale problems using novel and traditional large-scale data sources. MDI's strategic partnerships promote community and innovation across the health, social, computer, and data sciences.

MDI draws on expertise from across Georgetown's campus and beyond, including the social, natural, and computer science departments, and through strategic partnerships with organizations like the Beeck Center for Social Impact and Innovation, Lawrence Livermore Laboratories, the Institute for Social Research at University of Michigan, and the Pew Charitable Trust. The U.S. Census Bureau has designated MDI a Federal Statistical Research Data Center, one of only 23 in the nation.

MDI can support efforts to secure personally identifiable information and other sensitive data using industry-standard software encryption. We are also in the process of testing hardware-based encryption efforts and are happy to discuss other efforts we can take to preserve and protect all confidential data sources.

MDI has and utilizes a directory of standards and best practices for data linkage and research use. We have a network of centers and organizations who are facilitating data access for research and evaluation in housing, health, human services, workforce, justice, and education that meet and exceed federal and local laws and regulations. In addition, MDI is exploring methods to protect the interests and privacy of data subjects, including studies of ethical guidelines and statistical disclosure methods to reduce risks of re-identification.

HOWARD UNIVERSITY | School of Education

Expertise in Data Security

Each school and college at Howard University is required to have a Data Steward who is accountable for who has access to Institutional Data. The Data Manager is the data steward for the School of Education. Physical security of non-public sensitive data is maintained with badge entry. Access to such areas is granted only those employees who have a business need to know the sensitive information. Paper documents containing sensitive information are kept in locked office file cabinets or rooms. Only authorized employees have access to those spaces. Storage areas holding paper documents containing non-public information are always secure. No paper documents containing sensitive information may be removed from campus. Paper documents that contain personal, non-public information are shredded or securely destroyed at the time of disposal.

Howard University is required to meet the data protection standards in the Health Insurance Portability and Accountability Act (HIPAA) and the Family Education Rights and Privacy Act (FERPA). Access to data via the University's computer information systems is limited to individuals who have a business reason to know such information. Each employee is assigned a username and password, and multi-factor authentication is required. Databases containing non-public data are available only to School of Education employees with the business need to access the information. ZixCorp software is used for email encryption. Sensitive information on a system that is connected to the Internet is protected by a firewall. All non-public data stored on laptops or other portable devices are encrypted.

EXPERTISE IN DATA SECURITY AT MATHEMATICA

Protecting the confidentiality of sensitive data is of vital concern to DC and a cornerstone of Mathematica's work. As frequent collectors and users of data from the federal government, and state and local education agencies, we follow federal standards for using, protecting, processing, and storing data. Mathematica is highly experienced in and maintains a strong commitment to protecting the security and confidentiality of data that clients entrust to us. Our corporate security team develops, maintains, and regularly updates Mathematica's security policies, procedures, and technical safeguards. These policies are consistent with the Privacy Act, the Federal Information Security Management Act, the Family Educational Rights and Privacy Act, the Health Insurance Portability and Accountability Act, Office of Management and Budget memoranda regarding data security and privacy, and National Institute of Standards and Technology security standards.

Mathematica best practice: The project security checklist

Every project at Mathematica must complete a project security checklist within one month of the contract award date. We use the checklist throughout the life of the project to identify (1) project data sources and uses, (2) security and privacy requirements that apply to the project, and (3) appropriate procedural and technical safeguards. The checklist informs staff of the specific procedures and tools they should use to carry out project tasks. The project lead and others review the checklist annually and as needed during monthly internal project reviews and update the checklist as needed. The corporate security team reviews the initial and updated versions of the checklist to ensure that the project complies with contract security requirements and modifications.

For the DC RPP, Mathematica will draw upon our experience implementing security controls on other projects. Mathematica researchers and programmers are experienced in carefully adhering to these security controls, which we document in a detailed security manual. The Mathematica lead for a DC RPP project will ensure that all project staff sign a confidentiality agreement and receive annual security awareness training. Such training includes reviewing Mathematica's corporate computing rules of behavior. Our approach integrates security best practices into systems design, architecture, and operations. This approach includes securing data on ingress; securely storing, transforming, and loading data; and securely disseminating data as needed. More specifically, the computing assets that will comprise the system architecture for DC RPP projects will safeguard personal health information, personally identifiable information, and other confidential project information in a manner consistent with National Institute of Standards and Technology standards. Mathematica secures individually identifiable and other sensitive project information and strictly controls access to sensitive information on a need-to-know and least-privilege basis. In addition, data are encrypted in transit and at rest using Federal Information Processing Standard 140-2 compliant cryptographic modules. We securely dispose of all data according to our contractual and data use agreement obligations.

Trinity

Expertise in Data Security

Trinity Washington University protects personally identifiable information or other sensitive data, including research data, through software-based encryption. In accordance with industry best practices, data at rest and in flight is encrypted via use of RSA 2048 bit key pairs and where applicable, AES 256 bit symmetric keys.

In regards to access and administrative rights, Trinity employs the principle of least privilege whereby employees are assigned only those administrative rights and permissions needed to do their job.

COLLEGE OF EDUCATION

Expertise in Data Security

The University of Maryland has a sophisticated infrastructure to house a variety of data, both restricted and unrestricted, and can obtain specific requirements through the Division of Information Technology (https://it.umd.edu/). In terms of data management and analytics, depending on the project, the Measurement, Statistics, and Evaluation program has a server, referred to as *Biago*, which contains over two dozen different statistical and psychometric modeling software packages, to support the statistical simulations and application development to be undertaken in the project. *Biago* is a Dell Blade server with 256 GB of RAM. Specific software includes BILOG, Facets, flexMIRT, IRTPRO, Mplus, Parscale, Winsteps, HLM, JAGS, Lisrel, R, OpenBugs, Netica, SAS, STAN, and Stata. For most projects, the main infrastructure for storing the feasibility/stakeholder evaluation data that will be the online Box system, which allows for encrypted transfer and storage of up to 50GB of data per individual. This online resource meets guidelines for secure storage and analysis of sensitive data, as described by the UMD Institutional Review Board. For other applications needing higher security, a separate server can be utilized as necessary.

Expertise in Data Security

EdPolicyWorks has substantial experience in housing and using secure data. Most of the data EPW researchers employ come with data use agreements which stipulate the security measures that must be in place to use and store the data. In some cases, e.g., restricted access NCES data, these data require a secure data room only accessible by identified researchers with strict stipulations on the use of data and the resulting research products. In other cases, e.g., state policy partners, data must be stored on secure servers or secure computers in locked offices by approved personnel. EPW researchers currently have access to millions of observations that are stored on secure facilities. When requested data is stored in encrypted formats. We have never had a breach of security.

Appendix I

Validity, Data Quality, and Independence (3.1.6)

This appendix includes, as applicable, descriptions of Collaborative members' processes to ensure the validity and quality of research methods and outcomes; standard quality review processes applied to the creation of research products and deliverables; processes in place to ensure accountability and transparency in all work and independence with regard to funders, the public, and government entities; and Institutional Review Board processes.

URBAN INSTITUTE · ELEVATE · THE · DEBATE

Urban Institute Quality Assurance Process

This document provides quality assurance guidelines for Urban Institute research projects under our quality assurance process. Before a project begins, a quality assurance team will be identified and assigned to the project. This team will be composed of individuals with the computational and methodological expertise to guide and assist the quality assurance process.

The quality assurance process is divided into three main phases, which are described in the table below. The process kicks off with a pre-project meeting between the research team and quality assurance team before the beginning of phase I to discuss expectations. Each quality assurance phase ends with a checkpoint that will help the research team to self-assess whether they are successfully addressed the data quality considerations of the current phase and are ready to move on to the next phase. A meeting will also be held after the completion of each phase to review the checkpoint and discuss the expectations for the coming phase. After the final phase, a post-project debrief will be held to collect feedback about the quality assurance process and discuss lessons learned.

	Steps	**Checkpoint**	**Meeting**
Phase I: Research Design	Step 1: Understand Goal and Task Assigned Step 2: Understand Population Steps 3: Identify the Right Data and Methods	• Successfully complete pre-analysis plan at the end of Phase I • Checkpoint should be met before analysis begins	• Pre-Project Meeting before beginning Phase I • Pre-Analysis Meeting at the end of Phase I to review data acquisition & (pre-)analysis plan and discuss Phase II
Phase II: Analysis	Step 4: Acquire, Verify, Reconcile, and Prepare Data for Analysis Step 5: Develop Code and Run Analysis Step 6: Examine Results for Plausibility Step 7: Have Independent Review of	• Quality assurance monitoring of data collection and/or acquisition • Successfully complete code review at the end of Phase II • Checkpoint should be met before writing begins	• Pre-Publication Meeting at the end of Phase II to discuss results of code review and Phase III

	Code and Replicate Results		
Phase III: Writing and Dissemination	Step 8: Provide Output (Written and Verbal)	• Successfully complete publication data quality review at the end of Phase III • Checkpoint should be met before publication	• Post-Project Debrief following the completion of Phase III to discuss lessons-learned and collect feedback

Meeting: Introduce quality assurance team to project and discuss overarching quality assurance process and upcoming steps during phase I

Phase I

Step 1: Understand Goal and Task Assigned

During this step, the research team should develop a firm understanding of the goal of the project as well as other relevant work that has been previously completed within and outside Urban.

- Define the goal of the project and its research questions
- Define the key target audience(s) for the research and the specific product(s) (blog post, brief, fact sheet, report, etc.) that will be produced to reach the audience(s). This process should be completed in collaboration with Communications experts as part of the intake process.
- Communicate to all members of the research team the connections between research question(s)/goal(s), the research design, the data and analytic approach, and how the analysis will inform/address the research questions.
- Conduct a literature review (if possible given constraints)
- Identify relevant assets at Urban: experts on the subject matter, previous Urban publications, experts of relevant datasets, and existing code and tools (see step 3 for concrete suggestions)
- Ensure relevant experts have sufficient budget to complete review for steps 1 through 8.

Step 2: Understand Population

During this step, the research team will define the population of interest and relevant subpopulations (if applicable) and outline the summary statistics that will be run on the population/subpopulations.

- Determine the "population of inference" in terms of geography, time, units (e.g., people houses, establishments) and the "unit of analysis"
- Define relevant subpopulations (if applicable)
- Identify the statistical tests or summary tabulations to run on the target population (before looking at the data, if possible)

Steps 3: Identify the Right Data and Methods

- Identify data sources for data acquisition and analysis. As you consider each potential dataset for use and/or primary data to be collected, address the following:
    - the population covered by the dataset is the same as the target population for the study
    - the definitions of the variables in the dataset allow you to run the statistical tests and summary tabulations needed
    - the quality of the data for the key columns—what proportion of the observations are missing/imputed, do you have sufficient documentation to accurately understand the data, do the values of the columns align with what you expect from the documentation, etc.
    - is there a sufficient number of observations for each subpopulation of interest to enable us to perform the subpopulation analysis
    - if primary data are being collected, consider the methods (e.g., surveys, observations) and modes (e.g., self-administered vs. interviewer administered; web vs. telephone) for securing data of appropriate quality for analysis
- If primary data are being collected, there should be a plan that collects appropriate measures from the right source (e.g., person, organization, including sampling and recruitment plans) and using methods that promote high-quality data production.
- Identify whether the proposed method is adequate for answering the research questions.
- Complete an IRB application if necessary.
- Satisfy any restrictions associated with accessing the data (i.e., secure drive, confidentiality pledge, data license).
- Document when (date and data version number, if applicable), from where, and how you accessed the raw data file so someone could identify the version of the data you used for analysis and replicate your steps for accessing the raw data.
- Read the README and/or any documentation that is provided with the raw data.
- Identify the analytic variables and recoded variables that you will create from the raw data. Document the definitions of these variables and how they will be produced.
- Identify the diagnostics you can execute to verify that you are accessing and using the data correctly. This may include comparing summary statistics against those provided in the data documentation (if applicable), published tables/statistics from the data provider, or published tables/statistics from a trusted second party.
- Identify relevant assets at Urban: Urban data catalog, experts at Urban with previous experience using datasets, or the statistical methods group for primary data collection.
- Complete an analysis plan. As a best practice, whenever possible, we recommend pre-registering a full "pre-analysis plan" with the appropriate repository. As part of this process, we recommend projects write up their analysis plan for internal review, but submission to the repository is not required:
    - American Economic Association
    - Open Science Framework
- Urban maintains a list of resources as part of our internal data quality training, including:
    - Questions to ask your supervisor
    - Best practices checklist
    - Tools and resources

- If the project requires access to confidential data, you may want to think of the right code review person at this point, with help from the data quality team, to ensure they have access later on

Checkpoint: Complete (pre-)analysis plan

- In the review, we will be checking to ensure that a pre-analysis plan is written, and that it includes how you plan to address, or have addressed (for earlier steps), each of the 8 steps in the research process outlined here. Help is available from the data quality team to complete this step, as well as in the Urban Institute Data Quality guide.

Meeting: Review pre-analysis plan and discuss upcoming Phase II steps in the pre-analysis meeting, to be completed before data analysis begins.

Phase II

Step 4: Acquire, Verify, Reconcile, and Prepare Data for Analysis

- If primary data are being collected, implement the quality assurance monitoring plan to ensure data capture of appropriate quality.
- Identify relevant assets at Urban regarding code and programming: relevant code from previous projects, Urban or otherwise. Be sure to investigate previous code for any potential quality issues, or speak directly with the researcher who programmed it if possible.
- Eliminate manual edits to the data and ensure you document how the data were collected and provided to you, including any edits you made to the data.
    - Document the source of the information, including the URL, the date of the access, and specific metadata about the vintage. Oftentimes, this can be accomplished programmatically.
    - Make edits in code, not by hand.
- Verify that the original dataset is not corrupted. For example, text is replaced with odd unix characters, or a cell with a 10 in the original data online has an N/A in your data.
- Run basic diagnostic tests on the data matched against trusted publications, data, or other sources. Note, the match may not be exact because of statistical disclosure limitation like swapping.
    - Codebook:
    - Documentation:
    - Published tables or reports:
    - Other 2nd-party analyses:
- Review and identify how to address missing data codes, unexpected values (such as negatives for positive integer columns, decimals for integer values) and outliers.
    - Reference common practices for addressing missing data and outliers in respected literature using this data source.
- Resolve common errors by reviewing the list of "Common Mistakes During Testing."
- Conduct any additional checks according to the Promise Neighborhoods Introduction to Data Quality.

Step 5: Develop Code and Run Analysis

- When possible, all code, data, and code output should go in the same directory instead of being spread across multiple directories. This directory should be organized such that it can be moved anywhere on a computer, or to another computer, without affecting the functionality of its contents.
- Include a README in your project directory that describes how scripts and data relate (this is the same file that you can commit on GitHub and read easily). A README is a file that briefly describes the purpose and contents of a directory. Think of a README as the title page to your analysis. This file should contain a project title, an introduction/motivation, a list of the important contents in the directory and how they relate, any software requirements, and steps to reproduce the analysis (e.g., the order to run different scripts). It can also include a list of contributors and a license. If possible, this file should be a plain text file with the extension .txt or .md, though formats like Word may be easier for teams to implement.
- Use version control such as Git or GitHub (if necessary, you may use Box for documents but it is not advised for code) to store the master version of the code and possibly the data (file size permitting).
    - Version control is the management and documentation of changes to files. Good version control organizes and documents work so changes are recorded, communicated, and understood. It also structures collaboration so concurrent or asynchronous work is so easy it's common.
- Adopt a consistent and clear coding style. A few rules to keep in mind regardless of your programming language:
    1. Use descriptive nouns for variable and object names. Use descriptive verbs for function and macro names.
    2. Assign names to assumptions instead of hardcoding assumptions. For example, create a named object for a deflator and multiply by the object instead of multiplying by the unnamed number.
    3. Use hyphens and/or underscores in file names instead of spaces.
    4. Don't Repeat Yourself (DRY). Create a documented function or macro for any operation that occurs more than once. Assign frequently used scalars, vectors, or matrixes to named objects. This will reduce errors and require changing the code in one place instead of many places.
    5. Write functions and macros that are referentially transparent. This means a function or macro always generates the same output for a given inputs.
    6. Indent your code. Consistently.
- There are many more suggestions and rules that are language specific. Here are a few suggested style guides:
    - SAS
    - R
    - Stata
    - Python
- Add comments or use literate statistical programming tools like R Markdown, so others can easily inspect the work. Code should communicate the "what" of a script and comments should communicate the "why" of a script. Comments are essential but can be overused (e.g., commenting out code can create excess clutter and is not typically necessary when using systems that allow you to navigate to past versions of code, like

GitHub). Comments should be made above the relevant line of code. A common mistake is to change code without updating comments. Similarly, all functions should include a "function header," or a longer comment that explains the function purpose, inputs (the parameters the function expects and the data types of those parameters), and outputs (the object(s) the function returns and the data type(s)).

- For unfamiliar operations, test (on a small representative sample) that they operate as expected, and manually confirm the results of the code. The first ten rows of a dataset are rarely representative, so it is important to consider edge cases with adequate variation, extreme values, and uncommon values.
- Identify and write tests for analytic or recoded variables before constructing the variables and run diagnostics (i.e., cross tabs, scatter plots) to ensure the new variables were created correctly.
- Write tests that loudly fail and stop the analysis. It is too easy to stop looking at tests buried in logs. One way to consider organizing your tests is to write a test suite that you will run every time you update the code.
- Document the hardware and software, packages, and package versions that were used in the analysis. Many statistical packages have a function or macro that will document all hardware and software used in a session.

Step 6: Examine Results for Plausibility

- Implement simple face validity checks:
    - Do subpopulation totals exceed population totals?
    - Do numeric values exceed plausible or practical ranges?
    - Are distributions of categorical variables sensible?
- Implement "intuition tests," such as checking with experts or outside resources to ensure reported values are within their feasible range. Compare results of the time period of interest with adjacent time periods.
- Check that filters in the analysis were implemented correctly. Does the number of observations in the filtered data match tabs of the unfiltered data by the conditions of the filter?
- Check that joins in the analysis were implemented correctly. Join on the opposite conditions (or use anti joins) to examine the complement of the join. Should the anti join return any values at all?
- Benchmark your results against official statistics. If official statistics haven't been published, benchmark your results against statistics published in other sources:
    - Codebook
    - Documentation
    - Published tables or reports
    - Other 2nd-party analyses

Step 7: Have Independent Review of Code and Replication of Results

Code review is the process of soliciting outside feedback about the code used in an analysis. Mistakes and errors happen. Code reviews can be used to limit the chance of a mistake going unnoticed, ensure the validity of programs, and check the quality of documentation and comments. The two biggest challenges to code review are scarce resources and deadlines. These two challenges are also the biggest justifications for code review. Catching problems earlier,

particularly during analysis and before writing, can save resources and limit last second reruns and rewrites before deadlines.

- The best person to review code is 1) not already a contributor on the project (if possible), 2) knows the language used in the program, and 3) has some familiarity with the data used. In some cases, this may require more than one person. The reviewer should approach their responsibility with trust in the coder's intentions and skepticism of their code. The Urban data quality team can help in identifying the right person or people for a code review for the project as part of the pre-analysis meeting.
- When asking someone to review code, offer as little guidance as possible. If the README, directory, comments, and documentation are clear, then the reviewer should be able to understand without any assistance.
- That said, we recommend providing context for the project that might be helpful in catching errors. For example, if performing a ranking, helpful context is that a lower rank is defined as bad, and a higher rank is good. Meeting with the code reviewer and providing past papers and your current draft for reference is very helpful in this process.
- When preparing your analysis for a code review, we recommend you follow best practices used by top journal editors, such as matching the data and titles of output to your paper/product, having a single do-it-all master file that runs all your code and produces all your output for the product (and not all your tests, just for the paper/product), and citing data provenance/date (if your code does not pull the data directly). See the AEA Data and Code Availability Policy for more tips. This includes any code run by other teams at Urban, including the Communications or Technology and Data Science teams.
- Steps of Code Review:
    - **Step 1:** If possible, the reviewer should rerun the analysis on a different computer to see if the results are computationally reproducible. This is a good step to see if dependencies, sequencing, and computations are well managed. If the data are inaccessible to the reviewer because of confidentiality, consider making a dummy dataset with the same schema and characteristics as the confidential data.
    - **Step 2:** Check the code and logs for clarity, style, errors, warnings messages, and unpredictable results.
    - **Step 3:** Review the code against the pre-analysis plan and steps 4 to 6 in the data quality process to ensure that appropriate analysis and quality checks are in place as planned.
    - **Step 4:** The reviewer should outline any gaps in the documentation, program, or computational concerns to the analyst. Tools like GitHub allow for easy line-by-line review.
    - **Step 5:** The analyst should remedy any mistakes or omissions identified in steps 1 through 4. The analyst should return the fixed code to the reviewer to ensure that the issues from steps 1 through 4 have been adequately resolved.
- The above workflow isn't realistic for long-running projects like microsimulation models or massive data projects that require expensive cloud computing. There are often too many lines of code, or the costs of duplicating the analysis are too high. In this situation, diffs, or line-by-line logs of how code has changed, are invaluable. Proper version control becomes more essential to code review as the number of lines of code in a project increases.

Checkpoint: Complete code review.

Meeting: Review code review process and discuss upcoming pre-publication process in Phase III.

Phase III

Step 8: Provide Output (Written and Verbal)

- Properly annotate the sources of any graphics or tables, and provide notes to clarify the analysis, if necessary (so that the product "stands on its own"—in other words, if someone shared the graphic or table out of context without any of the associated text, the viewer should still be able to understand the key points correctly).
- To the extent possible, use an automated, and not manual, process to produce any graphics, tables, and statistics.
- The code reviewer should work with the research team as revisions are made to avoid last-minute errors or changes.
- The publication should be reviewed by a methodological or technical reviewer prior to publication, who should be identified early on in the project analysis plan and be kept up to date on potential deadlines to assure availability. This reviewer should be someone who was not involved in the research process or code review, if possible. This person should work closely with the code reviewer to ensure the code matches the final results and expresses those results accurately. The data quality team can help identify the right person for this task. The methodological reviewer will check for the following, at a minimum:
    - Graphics and tables are properly annotated and provide the data source.
    - The visual encodings in the graphics represent the underlying data (e.g., if a bar in a bar plot is taller than the other bars, it should represent a number that is higher than the other bars).
    - All numbers presented in graphics, tables, and statistics can be reproduced in the data analysis.
    - The language in the report accurately represents the data (e.g., is a "vast majority" backed up by the data, does the language account for margins of error, etc.).
    - The plots are visually consistent (e.g., if you are producing plots of factor variables, are the same factors always depicted as the same colors).
    - The graphics and tables meet the *"stand alone" test*. They can be fully understood on their own (e.g., if copied and pasted into an article by a journalist) without having to refer to the text.
    - A clear *statement of limitations* appears, including if necessary as a footnote in all tables and graphs.
- As a best practice, share and reference useful data on the Urban Institute Data Catalog.
- As a best practice, share and document your code on an Urban Institute GitHub repository.
- Be sure to include software packages, libraries, or repositories in your work citations and references list.
- The data quality team will assist in identifying the appropriate reviewer for this task. Appropriate lead time is required.

Checkpoint: Complete data quality review for all publications.

Meeting: Debrief to discuss the data quality assurance process, share lessons learned, and provide feedback on the process. The data quality team will general feedback as well as best practices and lessons learned. Research teams and reviewers are encouraged to share any documents used as part of this process if they think they would be helpful templates.

URBAN INSTITUTE · ELEVATE · THE · DEBATE

Urban Institute Center on Education Data and Policy Code Review Check June 2019

Researcher: Before Code Review

Prepare Your Code

- Build a file, or set of files in the same folder, that takes raw data from the folder, cleans it, and outputs only the tables and figures that are in your text. Delineate the "building" and "analysis" sections of your code (e.g., by including them in separate files).
- Consider starting your code with basic information, including date of last edit, programmer name(s), source file(s), and code purpose.
- When possible, use commands like *assert* to ensure that values make sense and that key variables are not missing (e.g., *assert* percentage_1+percentage_2=100%).
- Clearly label each analysis with comments on where it is in the text (e.g., "Table 1"; "% of Low-Income Students, Paragraph 2 on Page 5").
- Review your own code, looking specifically for the following issues:
    - **Missing Values**
        - Check for how missing values and zeros are coded.
        - Cap maximum values at x < . (do not use x != .).
    - **Merge**
        - Ensure that merges are matched appropriately (e.g., look at names in using and master datasets).
        - If a m:1 or 1:m merge, double-check the appropriateness of merge (e.g., schools to state-level data) and if consolidation may be needed. Merges using m:m should be avoided except in unusual circumstances.
    - **DRY (Do not Repeat Yourself) Issues**
        - Ensure that variables are coded only once, preferably at the top of your code.
        - Use reusable or modular functions, when appropriate.
    - **Force, Etc.**
        - Check that the use of *force*, *capture*, or similar shortcuts are the most valid approach for your data (e.g., consider cleaning strings of "Null" or "NA" values manually before converting to a numeric variable).

Gather Data Documentation

- Pull together codebook(s), datafiles, and do-files in an easy-to-access place for the code reviewer.
- If data manipulation is done outside the code (e.g., in Excel), make a note in the text with the relevant file path.

Pick Your Reviewer

- Pick someone who is able to access your dataset (e.g., has NCES license if working with restricted data).
- If possible, pick a researcher who has familiarity with your dataset. If this is not possible, consider also having a peer review of the text and methodology from an outside reviewer who has worked with the data before.

Identify Other Data/Papers on Topic

- If possible, identify other datasets or analyses that use the same data or assess the same topic and have similar (or dissimilar) results as the main findings of your paper.
- Highlight these cross-comparisons with a quick bullet-point list (e.g., "We found a national default rate of 7 percent for Parent PLUS loans—a paper looking only at Parent PLUS in Texas found a similar default rate, of 8.6 percent, over a comparable time frame.").

Provide Draft Text and Methodology

- Send code reviewer the current draft versions of the text, methodology, and Excel charts/tables, as appropriate.
- Flag any code that you have particular concern about or that diverge from previous findings.

Provide Project Charge Code

- Send your code reviewer the project code, along with an estimate of the number of hours needed to review and any relevant deadlines.

Code Reviewer: During Code Review

Read the Draft Product

- Read the draft text and methodology over before looking at the code.
- Note any places where there are unanswered questions on methodology or data.
- Check figures and tables for consistency as appropriate (e.g., percentages add to 100%; *n*-sizes are consistent, dollars were inflated over time).
- Check that data cited in the text match the figures (or describes source).

Review Researcher's Documentation

- Review provided cross-comparisons, if available, and skim codebook for basic familiarity of the dataset structure.

Run Code

- Run the code once to:
    - Ensure that you have access to all needed files.
    - Document sample size at the beginning and end of code.
    - Ensure that the code runs without erroring out.
    - Check for use of commands you're unfamiliar with.

Talk with Researcher

- Talk quickly with the researcher to ask any questions you have before beginning the review.
- Request any other data or information you need.

Check Data Cleaning/Building

- Review the data cleaning/building code, looking for:
    - **Variable Construction**
        - What is the unit of analysis (e.g., individuals, households)? Is it consistent throughout the dataset?
        - Are variables labeled correctly (e.g., is University of Michigan coded as a four-year public school)?
        - Are new variables what they say they are (look at the codebooks)?
        - How are missing data coded?
        - Use *tab*, *sum*, or *codebook* to review key variables—are the values what you would expect? Is it okay if minimums are below zero?
        - Is the code DRY (Do not Repeat Yourself)? Is there any place where a variable is rebuilt or changed later on?
        - If using files from different years or jurisdictions, are variable names aligned appropriately and any changes in variable coding accounted for?
    - **Simple Operations**
        - Check whether simple operations like addition/subtraction/division exclude observations with missing values. *egen* may be needed.
    - **Data Reshaping**
        - Look at *merges* and *appends*—do the data appear to be matched appropriately? Are there identical non-ID variables in both datasets (the master file will update them)? How are nonmatching data handled/ dropped? Is *update*, *replace*, or *force* specified?
        - Does the researcher subset the data at all? Permanently or temporarily?
        - Are *preserve* commands followed by *restore*?

- Are missing data removed before a *collapse* function?
- **Imputations**
  - Are data imputed? If so, how? Are there flags for imputed observations?
  - Do imputed data analyses look different than nonimputed data analyses?
- **Force/Capture/Quietly**
  - Run the code without these commands. Does an error occur? Does this error make sense (e.g., there's no data on pregnancy for a subset of male observations)?

Check Analysis

- Review the analysis data output, looking for:
  - **Tabs and Tables**
    - Add a ", *miss*" to tabs and tables. Does this change the results?
    - Do *if* statements include/exclude the correct population? You can check by using *sum* or *tab* on the relevant variables with the *if* statement.
    - Are weights used consistently? Conduct the analysis without weights, or with a different weight specification. Do the results vary substantially? If so, this should be a flag to verify that you're using the right weights. Using *rawsum* can also be a helpful check for number of observations.
    - Are tabs and tables copied over into Excel charts? If so, check that they were copied correctly. Using export commands like *outreg*, *estab*, *estpost*, or *export excel* are generally preferred to copying and pasting results by hand.
  - **Regressions**
    - If the author does lots of regressions, are they DRY?
      - Use a set of regressors in a local or global list?
      - Use loops to reduce amount of code? (Break loop if needed or run for just one instance).
      - Same population and same outcome variable across all regressions?
    - Check for consistency in regressions:
      - Does the number of observations vary substantially over the regressions? Should the number of observations be held constant across regressions for comparability?
      - Are weights used consistently? Conduct the analysis without weights or with a different weight specification. Do the results vary substantially? If so, this should be a flag to verify that you're using the right weights. Using *rawsum* can also be a helpful check for number of observations.

- Are coefficients in the direction that would be expected? Do controls make sense?
- Are appropriate standard errors used (e.g., robust, clustered, bootstrapped)? And are they calculated the same way across all regressions?

▪ Regression output:

- If regression results are exported, do they match up with what's in the terminal window?
- Are the *replace* and *modify* options specified correctly with exports?
- Are odds ratios, marginal effects, or other transformed results reported correctly?

Check Data Figures and Tables

- Review code and charts in Excel
  - Check functions and charts by clicking on them and ensuring that they refer to the correct cells. For columns of functions, check the top cell and a middle or lower cell to ensure that the function copied correctly (fix with $).
  - Double-check list of cells in functions: AVERAGE(A1, A2) is different than AVERAGE(A1, A2,).

Document Findings

- Make comments in a new saved version of the code with your corrections and questions. Use capital letters and sign with your initials so the researcher can easily find and review your comments.
- Write up other questions or concerns about the methodology or Excel figures in an email to the researcher(s).
- If large concerns exist, anticipate reviewing the code again.

Researcher: After Code Review

Review Code and Comments

- Look at the reviewer's comments and concerns.
- Reply to the reviewer's email with the steps you took to address any issues that were identified.
- In the code, make any necessary changes and carefully update your tables, figures, and text.

Address Further Questions

• If the code reviewer and the researcher are unsure about a question, the questions should be referred to a third researcher or to the Statistical Methods Group • If the code reviewer and the researcher continue to disagree, the questions should be referred to the center vice president via email.
After Code Review is Finished • Email the project administrator that the code check has been completed, which is a requirement before the R&R form will be signed by the center vice president.

Introduction to Data Quality

A Guide for Promise Neighborhoods on Collecting Reliable Information to Drive Programming and Measure Results

Peter Tatian

with Benny Docter and Macy Rainer

December 2019

Data quality has received much attention in the business community, but nonprofit service providers have not had the same tools and resources to address their needs for collecting, maintaining, and reporting high-quality data. This brief provides a basic overview of data quality management principles and practices that Promise Neighborhoods and other community-based initiatives can apply to their work. It also provides examples of data quality practices that Promise Neighborhoods have put in place. The target audiences are people who collect and manage data within their Promise Neighborhood (whether directly or through partner organizations) and those who need to use those data to make decisions, such as program staff and leadership.

Promise Neighborhoods is a federal initiative that aims to improve the educational and developmental outcomes of children and families in urban neighborhoods, rural areas, and tribal lands. Promise Neighborhood grantees are lead organizations that leverage grant funds administered by the US Department of Education with other resources to bring together schools, community residents, and other partners to plan and implement strategies to ensure children have the academic, family, and community supports they need to succeed in college and a career. Promise Neighborhoods target their efforts to the children and families who need them most, as they confront struggling schools, high unemployment, poor housing, persistent crime, and other complex problems often found in underresourced neighborhoods.[1]

Promise Neighborhoods rely on data for decisionmaking and for reporting to funders, partners, local leaders, and the community on the progress they are making toward 10 results. Crucial data for Promise Neighborhoods include required Government Performance Results Act indicators on academic achievement and family and community supports compiled from various sources, including

neighborhood and school surveys, school system records, and management and program data collected through direct interactions with service providers and people being served.[2] Without high-quality data, Promise Neighborhoods may not have the information they need to make good choices or to keep partners and allies accurately informed about their results. Promise Neighborhoods also receive and use data from partners who provide key services to children, youth, and families and need to ensure those data meet quality standards as well.

Data Quality Management

What Does Data Quality Management Mean?

Data quality management involves identifying the intended uses of data, creating criteria that are appropriate for identified data uses, and implementing practices to ensure data meet those standards. Data quality management aims to maximize the value of data for an organization. For a business, data quality management would maximize the value of data to generate profits. For a Promise Neighborhood, the focus would be on maximizing the value of data to make the best use of available resources to achieve successful outcomes aligned with the 10 Promise Neighborhood results or other community priorities.

There is no single, objective definition of what constitutes "good" data. Rather, data quality refers to an assessment of information based on its intended use and its suitability to serve that purpose. A Promise Neighborhood program may be based on an evidence model that requires a certain number of hours of classroom attendance. In that case, knowing total attendance time accurate within 15 minutes may be an appropriate standard. The data quality rules that Promise Neighborhoods put in place must reflect an understanding of why those data are being collected and how they are meant to help Promise Neighborhoods achieve their goals.

The rest of this brief describes basic data quality management activities, such as profiling data, creating data quality rules, and implementing data review processes. When building a data quality management process, it is important to understand that creating and maintaining high-quality data requires time and effort, and appropriate resources need to be devoted to these activities. Furthermore, everyone in the organization needs to be involved in ensuring data quality, which means many people will need to address data quality.

Allocating resources for data quality may be difficult in an environment where the priority is to channel funding to provide programming and services. Why spend money on data? In her article "How to Create a Business Case for Data Quality Improvement," Susan Moore describes five steps for building consensus with an organization for investing in data quality.[3] Her main advice is to not focus on data quality as an end in itself. Rather, proponents should emphasize how better data would benefit the organization's goals and mission. Furthermore, these arguments need to be framed in terms that are meaningful to both organizational leadership and staff, emphasizing their priorities and goals and how

better data will allow the Promise Neighborhood to serve the needs of children and families more effectively.

A key starting point is to create a data profile, which gives an initial overview of data quality and identifies problems. By elevating current challenges, data profiling can create an impetus for action to improve data quality.

Why Should Promise Neighborhoods Care about Data Quality?

The phrases "data driven" and "evidence based" are often used to describe how organizations try to conduct their operations in ways that produce measurable results. For nonprofits, data are starting to drive decisionmaking not only about what services should be in place but also in what manner, to whom, in which locations, and at what intensity. Data are also crucial in assessing whether interventions are achieving desired results—that is, whether programming is producing intended impacts. Being able to use data effectively is critical for nonprofits, including Promise Neighborhoods, that want to produce positive results for the people and families they serve.[4]

But while high-quality data can help Promise Neighborhoods improve outcomes, poor-quality data can do harm. As data become more influential in how people and organizations make decisions, the for-profit and nonprofit sectors increasingly recognize the damage that bad data can do. High-quality data can help Promise Neighborhoods in three significant ways.

- **Make good use of time and funding**. Having good data can save valuable resources. Even though federal grants for Promise Neighborhoods are large, resources are never sufficient to address all community needs. Those limited resources must be used effectively. Poor data can create wasted effort because time has to be spent correcting data before they can be used. The research and advisory company Gartner reported that organizations believe poor-quality business data are responsible for an average of $15 million a year in losses.[5] IBM estimated the annual cost of bad data to the US economy at $3.1 trillion.[6] Generally, it takes 10 times a much effort to complete a task using poor-quality data compared with using high-quality data because of the extra effort needed to find and correct errors.
- **Make better decisions.** Good data can lead to good decisions. The more that decisionmaking is data-driven, the greater the risk that poor-quality data will lead to wrong choices. If Promise Neighborhoods use data to decide which community needs should be addressed and which interventions are working, bad data can cause them to fail to identify crucial needs, direct resources to the wrong places or populations, or invest in the wrong solutions.
- **Enhance trust and credibility.** Promise Neighborhoods commit to transparency by using data to communicate with funders, partners, and the community about the efforts they are making, where they are expending resources, and what results they are achieving. Unrealiable information can harm the trust that people and institutions are willing to place in the Promise Neighborhood. The resulting loss of credibility could severely damage a Promise Neighborhood's ability to sustain and expand its work. But if a Promise Neighborhood is known

for being a good steward of data, it can build trust and credibility among partners and the community.

Roles and Responsibilities

Data quality should be the responsibility of everyone in the Promise Neighborhood who collect, enters, reports, or otherwise uses data. Nevertheless, assigning specific roles, responsibilities, and authorities for data quality is also advisable to ensure good data quality management practices are created and followed.

These roles include a *data quality manager*, someone in a leadership position who oversees quality efforts and helps create and communicate the Promise Neighborhood's data quality vision and goals. The Promise Neighborhood may also have one or more *data analysts*, who provide expertise in processing, checking, and reporting on data.[7] In collaboration with leadership, program staff, and partners, data quality managers and data analysts should conduct data profiling, develop rules and protocols for data quality, lead data quality reviews, and provide necessary training and support for people entering and processing data.

Managing Data Quality with Partners

Because of the wide range of services needed for a community-based, cradle-to-career pipeline, Promise Neighborhoods must work with partner organizations who can provide solutions the grantee backbone organization cannot. This makes data quality management even more challenging because, to some extent, Promise Neighborhoods must rely on partners to collect and report data without direct control or oversight.

Despite those challenges, Promise Neighborhoods should work with their partners to ensure quality standards are being met for all reported data. Along with other aspects of performance, Promise Neighborhoods should consider including data quality standards in their agreements with service providers. This brief can be a guide for how to structure those standards for partners, including establishing data quality rules and metrics. In addition, Promise Neighborhoods can use this brief as a training resource to help build the capacity of partners to collect and report high-quality data.

Data Profiling

Data profiling is the foundation of data quality management. The purpose of data profiling is to gain insights into current data quality and to provide a baseline against which future improvements can be measured. Data profiling consists of five steps.

- **Catalog all data being collected and used.** The first step is to document all data sources. Information should include the data source, when and how the data are collected, where the data are stored, what information is included, what the acceptable values for specific data elements are, how the data are used to meet business needs or goals, and which people are

responsible for collecting, entering, analyzing, and reporting on these data. This information is sometimes referred to as *metadata*, or data about the data. Metadata should be reviewed and updated regularly to make sure they accurately describe the data being collected.

- **Compare the data with their metadata.** After compiling the metadata, the next step is to see how well the actual data match the assumptions made about what those data should look like. For example, if the data are to be updated monthly, is that what actually happens? Or if a data field is supposed to consist of dates (month, day, and year), does it sometimes have other types of information? Or does it have irregular or improperly formatted values?
- **Run tests on the data.** An additional step is to examine basic summary statistics, such as averages or extreme values (highs and lows) to see if data values are consistent with expectations. For categorical data that have fixed values (e.g., "yes" or "no"), do any values fall outside accepted responses? Or are there inconsistencies in the data, such as numbers that do not add up to totals? The next section on data quality rules will give more examples of the kinds of problems for which tests can be run.
- **Report on data quality.** Reporting on data quality is an important part of data profiling. It is not sufficient for only a few people to know about data quality problems. Everyone needs to see them as an organizational challenge. The initial data quality report should include metrics of data quality and can be a baseline against which future improvements can be measured and reported on.
- **Repair the data.** Depending on the problems uncovered, the last step is to come up with a plan for addressing data issues. This should include not only fixing existing data and revising the metadata but taking actions to prevent the same problems from emerging in new data. To be effective, preventive actions need to be based on an understanding of the sources of data errors—that is, when, where, and how problems entered the data in the first place. Preventive steps include simplifying paper forms so they are easier to fill out and enter data from, as well as building in review steps so data problems are caught earlier.

These data profiling steps are the start of implementing a data quality management process, but they should be revisited periodically to track progress to ensure that improvements are made and that previously corrected problems do not reemerge.

BOX 1

Grantee Spotlight: Camden Promise Neighborhood

The Camden Promise Neighborhood in Camden, New Jersey, takes a proactive approach to data to ensure their accuracy. The team ensures the data are clean as they are input rather than fixing errors after they are in the system. Camden has a data management system that couples technical training with guidelines for navigating potential pitfalls. For example, the team tracks ages of clients and household members across various partner solutions for reporting purposes. If staff encounter unknown birthdays, they know to input 1/1/1900 so it is immediately obvious upon later review which clients need to be followed up with. Staff are trained to update data, including contact logs and DAP (data, assessment, plan) notes, on a consistent basis, and the team runs monthly reports as an additional check on their data quality assurance.

Data Quality Rules

Data profiling will reveal both underlying assumptions about and challenges in a Promise Neighborhood's data. Using this information, the Promise Neighborhood can develop data quality rules that define what high-quality data represent for the organization. Data quality rules specify the conditions the data must meet to be useful and can be grouped into five categories based on the issues they are meant to address: accuracy, consistency, completeness, integrity, and timeliness.[8] Data quality rules can also be used to set standards for partners on data reporting quality.

Accuracy

Accuracy refers to whether the data adequately reflect real conditions—that is, whether they are sufficiently correct and precise. Accuracy can have different meanings depending on how the data are used. At a basic level, accuracy is simply whether entered data match source records, such as intake forms or questionnaires. In table 1, information recorded on the paper student enrollment form has not been entered correctly into the database, making those data inaccurate. Although some incorrect entries may be obvious (for Maya), it may not be possible to detect other errors without comparing the database with the original source (for Sam).

TABLE 1

Accuracy Example

Enrollment form

Student	Age
Ana	12
Sam	11
Maya	9
Ken	10

Entered data

Student	Age
Ana	12
Sam	10
Maya	90
Ken	10

Further checking accuracy might involve verifying data against independent documentation. For example, one could look at birth certificates or school records to verify student ages. But it may also be possible to simply consult intake forms or talk with staff who know the families. Regardless of how it is done, incorrect entries must be resolved in the final data before they are used.

Consistency

Consistency means that data from the same or separate sources should agree with each other. For example, client home addresses should be the same across separate program records. Or, as in the example in table 2, the number of students listed in an after-school program participant roster should match the reported count of students enrolled. Inconsistent information can lead to confusion and may result in wrong or contradictory decisions being made based on which version of the data are being used. And even if data are *consistent*, they may still be *inaccurate*, so checking for consistency is not a replacement for verifying accuracy.

TABLE 2

Consistency Example

After-school tutoring program roster

Session day	Student
Monday	Ana
Monday	Sam
Monday	Maya
Monday	Ken
Tuesday	Dara
…	…

Total enrollment

Session day	Student
Monday	5
Tuesday	4
Wednesday	6
Thursday	8

Completeness

Completeness is based on whether a record is a full entry. Does it have enough information to draw necessary conclusions? Incomplete data may not useful, making them less reliable and less valuable. For example, clients in means-tested programs should have recent income amounts so proper eligibility can be determined. For measuring average daily attendance and chronic absenteeism, one needs to record whether all students are absent or present every school day. Gaps in the record, such as in table 3, affect the ability to calculate attendance measures accurately. In the example below, it would be particularly important to try to understand what happened on October 7, when attendance was not reported for any students, as well as why Ken's attendance is missing on so many days.

TABLE 3

Completeness Example

Student attendance records

	School Daily Attendance								
	10/1	**10/2**	**10/3**	**10/4**	**10/7**	**10/8**	**10/9**	**10/10**	**10/11**
Ana	Y	Y	Y	Y	?	Y	Y	Y	Y
Sam	Y	Y	?	Y	?	Y	Y	Y	Y
Maya	N	N	N	Y	?	Y	Y	Y	Y
Ken	?	?	N	Y	?	?	?	?	?

Note: Y = in school; N = not in school; ? = attendance not reported.

Integrity

Integrity is based on whether data are stored properly based on the database structure. The metadata should describe what type of information is expected in different data fields. Is a field expected to contain numbers or text? Phone numbers or dates? Coded values ("Y" and "N" for yes and no)? Deviations from expected values need to be noted, as they can cause interpretation problems. In the example in figure 4, a Promise Neighborhood partner has provided a roster for families enrolled in one of its programs, but phone numbers are inconsistently formatted, which makes it hard to identify invalid numbers, such as for the Fernandez family (no area code) and the Johnson family (too few digits). The field for Promise Neighborhood residency also contains different types of answers, which will make those data harder to query or summarize later.

TABLE 4

Integrity Example

Family roster

Family	**Home phone**	**Promise Neighborhood resident**
Allen	202-555-0100	Yes
Fernandez	555-0101	Yup
Wang	2025550102	N
Garcia	(202) 555-0103	7
Johnson	202-555-010	Not sure???
Lee	202x555x0105	Uh huh

Having good metadata can ensure data integrity because the metadata will describe what data values are allowable for different entries. Sharing the metadata or user guidelines for data entry can also help partners and people responsible for collecting and entering data understand what is expected.

In addition, many formal database systems support data validation rules for specific fields so data integrity can be enforced when information is entered. For example, a rule can be created requiring all phone numbers to have 10 digits and preventing users from entering something with fewer or more numbers. Sophisticated systems can automatically validate street addresses by comparing them against a known list.

For Promise Neighborhoods or partners who use spreadsheets to store data, however, enforcing data integrity can more challenging. By default, spreadsheets allow people to enter a mix of information (numbers and letters) into cells within the same column (figure 4). Nevertheless, spreadsheet programs such as Apache OpenOffice Calc, Google Sheets, and Microsoft Excel also support creating validation rules for entries like phone numbers and coded values or for checking ranges of values (minimums and maximums). The Promise Neighborhood data manager could set up and apply appropriate rules to data cells before making worksheets available to program staff for data entry.[9]

BOX 2

Grantee Spotlight: Hayward Promise Neighborhood

Hayward Promise Neighborhood in Hayward, California, collects data at the intervention level, noting student ID; full name; intervention type, date, and duration; and miscellaneous notes. These data are stored in an Excel spreadsheet, and the data team manually checks for errors, understanding that certain critical fields (e.g., eligibility status) require an additional check. But the team is transitioning to Salesforce next year to make data input and cleaning smoother. In the past, the neighborhood has had a designated group of "data champions" that created norms and culture around data. Because of staff rotations, the level of excitement and organization around data has fluctuated, but Hayward emphasizes that partners are responsible for accurate data entry and reporting.

Timeliness

Timeliness reflects how long it takes between when data are expected and when they are available for use. Promise Neighborhoods can establish quality rules for the time it should take for data to be entered into their systems. Rules might specify that intake data on clients should be entered and be available for reporting within one business day or that student attendance data should be uploaded weekly. Making sure data are accessible quickly is an important aspect of ensuring their value to the Promise Neighborhood and others.

Data Quality Metrics

For each of the data quality rules described above, it is possible to count the number of times rules are violated in a portion of data (e.g., data collected over the past three months) and report these as metrics of data quality. Quality metrics can be expressed in many ways to gain insights into the extent of data problems.

- **Errors in dataset.** Number of distinct data quality rule violations in all data fields and records. Measures the volume of data problems.
- **Errors in dataset by category.** Separate counts for accuracy, consistency, and completeness problems. Identifies the most common data errors, which can be used to prioritize corrective and preventive action.

- **Errors in dataset by quality rule.** Separate counts of problems by quality rule. Identifies problematic data fields.
- **Records with errors.** Number of data records (i.e., observations for individual people or families), with one or more data quality rule violations. Shows how many data records are compromised and how many are clean.

Data quality metrics should be related to potential impacts of poor-quality data on Promise Neighborhood programming and the neighborhood's ability to provide effective services and produce positive results for children and families. Metrics should be reviewed regularly to assess the current state of the data and compared over time to measure progress. The frequency of reviews should depend on how critical these data are to the program's success and how likely the data are to have problems.

Data quality metrics should be included in regular reporting to Promise Neighborhood leadership, staff, and partners to keep everyone informed about the state of data quality and to generate buy-in to the idea that data quality affects the entire organization and is everyone's responsibility. Promise Neighborhood partners should be able to conduct their own reviews and report on their data quality, and data quality review tasks and reporting can be written into partner agreements.

Data Quality Review

In addition to reporting on data quality metrics, a Promise Neighborhood can implement regular data quality review with staff. Such a process will create an opportunity for people to interact directly with the data, which can facilitate coming up with strategies for improving data quality. Promise Neighborhoods could also include data quality reviews as part of their regular work with partners to improve the quality of their data.

Friday Afternoon Measurement is an example of a data review process that can be incorporated into a Promise Neighborhood's regular practice.[10] The process can be used by managers and others whose work depends on data collection and consists of four steps.

- **Step 1.** Assemble the most recent 100 data records used or created for some part of the Promise Neighborhood's work. These might be survey data, intake data from staff or partners for a particular program, or administrative data from schools or other agencies. Focus on 10 to 15 critical data elements—that is, specific pieces of information that are most relied upon for operations, reporting, or evaluation. Put all these data in a spreadsheet or print them.
- **Step 2.** Convene a two-hour meeting with several people with knowledge of the data. These people should be familiar with how the data are collected or used and who must rely on these data for part of their work.
- **Step 3.** Working record by record, highlight obvious errors (e.g., misspelled names or incorrectly placed information) in a noticeable color, like red or orange. Do this quickly, spending no more than 30 seconds on a record.

- **Step 4.** Summarize the results by adding a "Perfect record" column to your spreadsheet. Put a "yes" in this column if no errors were found or "no" if any red or orange appears in a row. Add up the number of perfect records.

TABLE 5

Data Quality Review Example

Student attendance records

Rec #	Student	Date enrolled	Sessions Attended					Pre score	Post score	Perfect record?
			1	2	3	4	5			
1	A	1/2/19	Y	Y		Y	N	40		No
2	B			Y	Y	Y	Y	20	40	No
3	C	2/7/19	Y	N	N	N	Y	30	20	Yes
4	D	2/12/19	Y	Y	Y	Y	Y	2	50	No
5	E	2/12/19	Y	Y	Y	N	Y	60	80	Yes
...	...	...	...	...	...	...	...	...	...	...
100	ZZ	3/12/18	Y	Y	Y	Y	Y	50	50	No

Number of perfect records = 52.

Notes: Enrollment dates should be between 1/2/19 and 3/31/19. Pre and post scores should be between 40 and 100.

During data quality review, participants found several issues, including missing data (e.g., enrollment date, attendance, and post score in records 1 and 2) and invalid entries (e.g., pre score and date enrolled out of range in records 4 and 100, respectively). Only 52 records had no detectable data quality issues.

Participants could then discuss the results. Was the number of problems surprising? What challenges do these data quality issues create for the Promise Neighborhood in understanding the program's effectiveness and communicating the program's value? What are the likely causes of the data challenges? What strategies could prevent the most prevalent or damaging data problems from occurring? How could those strategies be tried and tested? Ideas generated during the session could then be tried out, and progress could be assessed in the next data quality review.

As with all Promise Neighborhood performance review and accountability processes, the focus of a data quality review should not be punitive. The review is meant to identify challenges and develop solutions to ensure high-quality data to achieve robust results for the people and communities Promise Neighborhoods serve.

Conclusion

Robust data quality management will help Promise Neighborhoods better serve their communities by providing more reliable information that can be used to more accurately direct resources, assess results, and improve performance. To be successful, data quality management must be supported by leadership and built into regular practice. Data profiling is an effective place to start, because it will help everyone understand what data quality challenges exist and what impact they may be having on neighborhood

results. Once data profiling is complete, Promise Neighborhoods can add other elements of data quality management, including data quality rules and metrics. Regular data quality reviews, whether monthly, bimonthly, or quarterly, can culminate in strategies for improving data that can be implemented and tested. By following these recommendations, Promise Neighborhoods should be able to produce measurable improvements in data quality.

Notes

[1] "Promise Neighborhoods Program," US Department of Education, Promise Neighborhoods, accessed December 14, 2019, https://promiseneighborhoods.ed.gov/background/promise-neighborhoods-program.

[2] For detailed information on Promise Neighborhood data collection and reporting expectations, see Comey et al. (2013).

[3] Susan Moore, "How to Create a Business Case for Data Quality Improvement," Gartner blog, June 19, 2018, https://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement/.

[4] Effective use of data is one of the core competencies identified in PNI (2014).

[5] Moore, "How to Create a Business Case."

[6] "The Four V's of Big Data," IBM Big Data and Analytics Hub, accessed December 14, 2019, https://www.ibmbigdatahub.com/infographic/four-vs-big-data.

[7] Mona Lebied, "The Ultimate Guide to Modern Data Quality Management (DQM) for an Effective Data Quality Control Driven by the Right Metrics," Datapine, June 28, 2018, https://www.datapine.com/blog/data-quality-management-and-metrics/.

[8] Adapted from Lebied, "The Ultimate Guide."

[9] For tips on setting up data validation rules in Apache OpenOffice Calc, see "Validating Cell Contents," Apache OpenOffice, accessed December 14, 2019, https://wiki.openoffice.org/wiki/Documentation/OOo3_User_Guides/Calc_Guide/Validating_cell_contents; for Google Sheets, see "G Suite Pro Tip: How to Create a Dropdown List in Google Sheets (and Pointers on Conditional Formatting)," Google Cloud, October 17, 2018, https://cloud.google.com/blog/products/g-suite/pro-tip-how-create-dropdown-list-google-sheets-and-pointers-conditional-formatting; for Microsoft Excel, see "Apply Data Validation to Cells," Microsoft, accessed December 14, 2019, https://support.office.com/en-us/article/apply-data-validation-to-cells-29fecbcc-d1b9-42c1-9d76-eff3ce5f7249.

[10] The creator of this process, Thomas C. Redman, chose the name because "many people set up these meetings on Friday afternoon, when the pace of work slows." For more information, see Thomas C. Redman, "Assess Whether You Have a Data Quality Problem," *Harvard Business Review*, July 28, 2016, https://hbr.org/2016/07/assess-whether-you-have-a-data-quality-problem.

References

Comey, Jennifer, Peter A. Tatian, Lesley Freiman, Mary K. Winkler, Christopher R. Hayes, Kaitlin Franks, and Reed Jordan. 2013. *Measuring Performance: A Guidance Document for Promise Neighborhoods on Collecting Data and Reporting Results*. Washington, DC: Urban Institute.

PNI (Promise Neighborhoods Institute). 2014. "A Developmental Pathway for Achieving Promise Neighborhoods Results." Oakland, CA: PolicyLink, PNI.

URBAN INSTITUTE · ELEVATE · THE · DEBATE

Urban Institute Funding Principles

The Urban Institute strives to meet the highest standards of **integrity** and **quality** in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. The organization's reputation for integrity and quality is our most valuable asset, which we seek to protect and enhance.

We are grateful to our funders and partners who make it possible for Urban to advance its mission. We believe that operating consistent with the values of independence, rigor, and transparency is essential to achieving our shared goals.

The principles described below are designed to help us advance our mission and the objectives of funders through consistent application of these values. This is a living document that will be regularly reviewed and updated as necessary and as we learn from experience.

Funding shall enable and not conflict with mission
The pursuit of essential support should not divert the Institute from its mission. Therefore, Urban will solicit and accept support only for activities that are consistent with its mission.

Adding to public understanding and sharing of ideas
Urban's mission of opening minds is best advanced by sharing the data, methods, findings, insights, and knowledge gained through our work. We add to the public's, practitioners', policymakers', and researchers' knowledge in order to improve outcomes and elevate debate. Therefore, we strive to disseminate widely the findings and insights from our work. In some instances, we also will work with funders to help them understand the implications of our analyses for their own work and decisionmaking.

Independence from funders
No funder shall determine research findings or the insights and recommendations of our experts. Urban will not accept any support that implies or requires endorsement of a position or product. We welcome comment, correction, and substantive engagement from all informed parties to enhance the quality of our work, but the board and management of the Institute will defend the independence of researchers and experts, even if funders disagree with their findings or conclusions.

Intellectual property
The Urban Institute retains rights in intellectual property produced during and after the funding period. After a piece of funded work is complete, we strive to communicate the insights and conclusions of our research through other formats, including essays, interactive features, data

visualizations, or blog posts, subject to reasonable constraints imposed to protect individual privacy or proprietary information. We provide funders with reproduction and distribution rights for research reports they have funded.

Expectation of researchers
As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.

Diversity of thought
We welcome and celebrate the diversity of our staff members and partners across many dimensions, including the range of academic disciplines and issue areas represented, the variety of research and analytic methods used, the breadth of modes of inquiry followed, and the unique experiences and perspectives that each employee brings. We believe diversity spurs innovation while improving the quality of our work. We are comfortable when different Urban Institute experts, examining the same or related questions, reach different conclusions, provided that they are transparent about their methods and the work meets our standards of quality.

Disclosure of funding sources
All contributions will be publicly acknowledged annually and the list posted on our website. We list funders in categories based on the amount of funding received for each year. Urban will consider granting anonymity to individual funders who request it, subject to evaluation of the best interests of the Institute and the funder with regard to transparency, reputational risk, and the limits of the law. Where particular products or events are made possible by a particular funder, the funding relationship shall be disclosed in connection with that product or event along with standard language affirming our research independence.

Freedom to decline support
Urban can, at any time and for any reason, reject or return support from an individual, corporation, or foundation at management's discretion.

Protection of the brand and reputation for quality and independence
Management will guard the appropriate use of the Institute's name, logo, and reputation. Funders may not use the Institute's brand or visual identity without advance permission for each specific use.

American University
School of Education

Research Quality and Independence

Research, scholarship, professional and creative activities conducted at American University and each of its divisions (including the School of Education) are to be conducted according to the highest ethical and professional standards. The University seeks to ensure integrity in the design, conduct and reporting of research results**.** All members of the AU community have an obligation to report in good faith suspected research misconduct. The AU Office on Research Integrity ensures that all research policies and procedures are of the highest standard. The AU Office of Research is currently directed by the Interim Vice Provost for Research, Dr. Sarah Irvine Belson, Executive Director of the Institute for Innovation in Education and SOE faculty member.

A framework for imparting the "best practices" associated with ethical and professional standards, and increasingly regarded as a critical component of scholarly and career development, is training in the Responsible Conduct of Research (RCR). The initial National Institutes of Health (NIH) policy on RCR was published in 1989. RCR training was mandated for graduate students and postdoctoral fellows and faculty funded by the NIH training grants and career awards. New standards for RCR training were issued by the NIH on January 24, 2010. RCR training is required for NIH Institutional Research Training Grant Awards, Individual Fellowship Awards, Career Development Awards (Institutional and Individual), Research Education Grants, Dissertation Research Grant Awards, and other grant awards with a training component and must meet more specific guidelines. As of January 4, 2010, RCR training is also required for undergraduates, graduate students, and postdoctoral fellows funded by the National Science Foundation (NSF). Beyond these regulations, RCR training is strongly encouraged for all AU faculty, staff, and students engaged in scholarly work, regardless of funding source or field of study.

RCR training for all AU faculty, staff and students includes the following topics:

- Research Misconduct (falsification, fabrication, plagiarism)
- Conflict of Interest and Conflict of Commitment
- Authorship and Publication
- Data Management and Data Ownership
- Mentor and Traineeship Responsibilities
- Peer Review
- Collaborative Research
- Protection of Human Subjects
- Welfare of Laboratory Animals

Other topics frequently considered in RCR training include ethical deliberation, whistle blowing, lab management, environmental and lab safety, intellectual property, national security and export control, research in international and intercultural environment, public diplomacy/policy.

All Human Subjects Research conducted by students, faculty, or staff of American University must receive approval from the American University IRB. The IRB has determined that most classroom research, many oral history projects, and some review of preexisting data will not require IRB approval. All IRB protocol application forms are submitted via Cayuse IRB.

Challenges to the validity of research methods**:** SOE faculty, like most AU full-time tenured, tenure-track and research faculty, submit research articles and proposals through refereed or peer-review processes. All of our participating faculty have extensively published research in top-tier journals and publications, with a high degree of scrutiny and evaluation of their research methods.

Research Quality and Independence

Internal Review Processes

To ensure the validity and quality of research methods and outcomes, Bellwether's research projects are designed and reviewed by staff with the appropriate content and methodological expertise. Different team members have clearly designated roles and responsibilities with regard to ensuring quality in our final products. Analysts are expected to adhere to data cleaning and management processes, including documenting steps in the data analysis process and annotating code for replicability. Project managers are responsible for ensuring that final deliverables undergo a data- and fact-checks prior to completion. Products are reviewed for strength of evidence base in supporting arguments made.

Handling Challenges and Variation in Data Quality

Bellwether's policy and evaluation team has extensive experience with data collection, cleaning, management, and analysis. Our processes and procedures are aimed at ensuring consistency in handling challenges and variations in data quality. For example, the data management protocols outline steps for evaluating and documenting source data. Staff are expected to document decisions made during data collection, cleaning, management, and analysis.

Quality Review Process

Bellwether has established processes to ensure quality of final deliverables. During the data check, a Bellwether team member not involved in the project checks numbers in the final deliverable against output from statistical software and provides feedback on the presentation of findings. The team has a Quality Review checklist to ensure consistency and quality in final reviews. This checklist includes items specific to organization and format, writing quality, and ensuring data sources and evidence are sufficient to support the argument made. Field-facing publications are reviewed by external experts who provide feedback and serve as an additional check on quality. All final deliverables, whether field- or client-facing, are reviewed by senior team leaders who provide feedback, which is addressed prior to submitting the final product.

Processes to Ensure Accountability, Transparency, and Independence

Integrity is one of Bellwether's core values and is reflected in our approaches to accountability, transparency, and independence. Bellwether team members retain absolute editorial control over the content of any published editorial work carrying their byline or the Bellwether brand. We approach all our work unencumbered, meaning we will never shape our opinions to suit a client's interests. Our diversity of clients and funders is another way we protect our editorial independence. Bellwether is a strictly non-lobbying/non-representative organization. We work for our clients but we do not represent them or speak on their behalf. At no time during any engagement will any Bellwether team member or contractor engage in any work that would require a lobbying disclosure pursuant to any federal, state, or local statute or regulation.

Bellwether always holds clients' proprietary information in confidence, but we disclose the existence of any formal relationship between Bellwether and any entity. To ensure transparency, all relevant client and organizational relationships are revealed in any published work, and all clients, past and present, are listed on our website.

Institutional Review Board

When projects require Institutional Review Board (IRB) review, Bellwether works with partner organizations and/or an external IRB to ensure the protection of human subjects. Key staff have completed human subjects ethics training.

RESEARCH QUALITY & INDEPENDENCE

The core values of the Brookings Institution are quality, independence, and impact. Producing work that is rigorously conducted, objective, and useful is among Brookings' highest priorities.

We employ both formal and informal processes to ensure the validity and quality of our work. All policy briefs and policy reports released by the Brookings Governance Studies program must be reviewed by at least one Brookings scholar with relevant expertise. In addition, these briefs and reports are reviewed by a research assistant and communications staff to fact-check the work and ensure it is clear, readable, and sound. As necessary (e.g., if a brief or report involves technical work outside the expertise of would-be reviewers), additional formal reviews are undertaken by a relevant expert either inside or outside of Brookings.

In addition to our formal review process, we seek out opportunities to obtain feedback from an assortment of audiences. This includes assessments of our methodological approaches and interpretation of results. For example, we regularly present work at academic conferences in order to obtain and respond to feedback prior to publication. We publish much of our work in academic journals, subjecting it to the journals' peer review processes. We also actively solicit feedback from researchers, policymakers/practitioners, and others who are willing and able to provide critical comments.

With respect to transparency and accountability, we make our research available free of charge to the public on the Brookings website, through newsletters, and via other outlets. We strive to be as clear and transparent as possible in describing our methods and analyses.

Research Independence

Brookings has adopted a Research Independence Policy that enshrines in a single document the requirement that its scholars be independent and act accordingly. Specifically, the Institution and its personnel, including all scholars, may not engage in activities—on behalf of Brookings or others—that run afoul of the IRS prohibition on a substantial part of the Institution's activity constituting attempts to influence legislation; that would require registration under the Lobbying Disclosure Act, the Foreign Agents Registration Act, or state and local lobbying disclosure laws; or that might otherwise call into question the independence and objectivity of the Institution's research. As part of this

commitment to safeguarding independence, Brookings has established a set of Donor Guidelines that govern all fundraising and donor engagement activities by resident and non-resident scholars, development officers, and other staff. Brookings will not accept gifts from donors who seek to undermine the independence of its scholars' research or otherwise to predetermine or influence recommendations. Brookings does not conduct proprietary research.

Brookings scholars and other personnel must not permit the interests of any third party—whether a political party, government (foreign or domestic), donor, or non-governmental organization—to adversely influence their methods or conclusions. Brookings scholars may seek input from a number of interested stakeholders but they ultimately have the final word on their research and other activities.

D.C. POLICY CENTER

Research Quality and Independence

The D.C. Policy Center's Education Policy Initiative team ensures that its research is high-quality and error-free through a series of internal and external review processes. For each study, the team first surveys available data and literature to inform a draft methodology, which is reviewed internally by other staff before an external review by the Education Policy Initiative Advisory Board (six local education experts with deep knowledge of the District and public school choice). After the analysis is completed, the team checks each finding and shares the draft report with lead education agencies, experts, and the Advisory Board for a final review.

Approach to handling challenges and variation in data quality
The Education Policy Initiative team is careful to examine data to ensure it can be used for analysis. These data checks include internal validation where appropriate and cross-referencing other datasets. In the case of an incomplete dataset that does not majorly affect the analysis, the team clearly notes which data are missing so that stakeholders know how to interpret findings.

Processes in place to ensure accountability and transparency in all work, and independence with regard to funders, the public, and government entities
The D.C. Policy Center's Education Policy Initiative values accountability and transparency in its work. With the public and government entities, the team consults with key stakeholders to develop its research agenda, and posts topics for upcoming reports on its website. Each report undergoes a rigorous review that includes local experts and education agency representatives with the dual goals of ensuring accurate use of publicly available datasets and providing lead time to government entities to interpret findings. The team also reports regularly to funders against predetermined metrics and shares reports in advance of publication. The Education Policy Initiative maintains independence in its research: at no point do funders, reviewers, or government agencies influence findings.

Whether your research outcomes have ever been challenged based on the validity of the methods
Our research outcomes have never been challenged based on the validity of the methods.

Internal review processes and/or protocols to ensure the validity and quality of research methods and outcomes

Internally, our SQL tools and Microsoft Power BI query trackers perform robust quality checks on the data we receive from school partners. This process ensures reliable data and improves analytical efficiencies. We also rigorously investigate outliers in the data to understand the true nature and authenticity of the outlier. EmpowerK12's extensive knowledge of school data and how it's collected across DC will be a helpful resource for all Collaborative members attempting to understand the validity of outliers.

Approach to handling challenges and variation in data quality

EmpowerK12 takes a proactive and collaborative approach to educational data quality in the District. Our proactive strategies include data manager trainings on compliance and data quality best practices, professional learning communities for newer and veteran school data personnel, and sharing of SQL code so schools can run their own data checks throughout the school year. Empowering all schools' staff with the background knowledge required to manage their own internal data validity, structure, and reliability is one of the primary ways our team plans to support the Collaborative's success.

Standard quality review process applied to the creation of research products and deliverables

The EmpowerK12 internal quality review process involves multiple members of the organization. First, the project owner seeks out at least two members of the organization, including at least one senior director, with the most appropriate expertise for the project and requests a quality review. Those reviewers independently markup and analyze the work, then come together to build a consensus on the feedback that should be given to the author. Finally, the reviewers and the project owner meet to discuss the feedback and determine next steps, publishing only research that meets our high standards and adequately details possible sources of analytical variance.

Whether your research outcomes have ever been challenged based on the validity of the methods

EmpowerK12 applies the latest small-scale scientific methodologies and statistical controls in all our analyses for education partners. When conducting value-added analysis, we aggregate results across dozens of algorithms to avoid potential bias and reduce error in the calculation. If possible, depending on the nature of the data, we post raw data as well as model specifications online with reports that allow external reviewers to replicate the findings from our work. The team takes these steps to build trust in the validity of research findings.

Research Quality and Independence

Georgetown University's Office of Research Oversight (ORO) was formed in November 2018 to reflect Georgetown's ongoing commitment to consolidate and harmonize research compliance activities across all campuses. ORO's mission is to facilitate basic science research and protect the safety and welfare of human subjects participating in research and clinical trials by ensuring that all activities are conducted in a legally and ethically compliant environment.

Regulatory affairs and the Institutional Review Board's (IRB) primary role is to safeguard the rights and welfare of all human subjects who participate in research studies conducted by Georgetown.

In compliance with Federal law and institutional policy, all research projects involving human subjects or human material must be reviewed and approved by the IRB. All biomedical, social and behavioral research projects conducted by the faculty, the staff and students of the University are subject to the Policies and Procedures of the Institutional Review Board.

The Georgetown University IRB has the authority to disapprove, modify, or approve protocols based upon consideration of human subject protection. It also requires progress reports from the investigators at least annually and oversees the conduct of the study.

The overall criteria for IRB approval are:

1. The risks to subjects are minimized as much as possible.
2. The risks to subjects are reasonable in relation to anticipated benefits.
3. The informed consent is adequate.
4. Where appropriate, the research plan makes provisions for the safety of the subjects during the data collection process.
5. Where appropriate, there are adequate provisions to protect the privacy of subjects and maintain the confidentiality of data.
6. Appropriate safeguards are included within the study to protect the rights and welfare of the vulnerable subjects.

HOWARD UNIVERSITY | School of Education

Research Quality and Independence

- Internal review processes and/or protocols to ensure the validity and quality of research methods and outcomes
  The university requires all faculty who conduct research to successfully complete human subjects training through the Collaborative Institutional Training Initiative (CITI) Program with a score of 80 or above. Additionally, students, whether conducting research on their own or working with faculty, must successfully complete CITI as well as the online and in-person Responsible Conduct of Research (RCR) courses. Passage of these courses help ensure studies are being conducted ethically in order to help produce valid results. The School of Education has numerous faculty who are experts measurement, research methods and statistics, psychometrics, and evaluation. These faculty are often sought out to provide guidance to others from both within and outside of the university. At least one of these faculty are included or consulted formally or informally about the validity (inclusive of reliability) of studies conducted by those in the School of Education.

- Approach to handling challenges and variation in data quality
  As faculty in a School of Education, we are quite familiar with challenges and variations in data quality. Specifically when working with schools in a district with high student mobility, faculty have learned they must be flexible and adaptive. While some data received or collected is high quality, other data may not be due to various reasons. Faculty use what is usable and have sometimes adjusted research questions and data analyses based on the data received. Researchers are also aware that all data is not usable so we accept this, try to collect additional data, and use this as a learning experience for the next study.

- Standard quality review process applied to the creation of research products and deliverables
  The first step of the review process occurs before research studies begin since studies must first receive IRB approval. The IRB's responsibility is the protection of human subjects. This process ensures participants are treated ethically and the researchers will have the adequate resources to conduct the research, among other things. Once studies are underway, as a highly collaborative school, faculty share their work or sections of their work to gain constructive feedback from others. This strengthens the work and provides an informal review process. The formal and informal process has worked well for School of Education faculty and students.

- Processes in place to ensure accountability and transparency in all work, and independence with regard to funders, the public, and government entities
  All research at Howard University is governed by the Office of Regulatory Research Compliance (ORRC), which operates under the Associate Provost for Research. The mission of the Office of Regulatory Research Compliance is to oversee research compliance, interface with faculty, staff, and students; and work collaboratively with various committees and units, to foster and safeguard the institution's culture of compliance with applicable Federal, State and Local Regulations, Tribal Laws, and Institutional Policy and Procedures. The ORRC assurance applies to the ethical conduct of research involving human subjects, animals, hazardous materials, use of recombinant materials (rDNA, and synthetic nucleic acids), research integrity, conflict of interest and export control.

  The University requires mandatory training for all researchers, offered through the Collaborative Institutional Training Initiative Program (CITI) and offers regular workshops that include ethics training, compliance workshops, and policy updates.

- Institutional Review Board process
  All research projects dealing with human subjects must be approved by Howard University's Institutional Review Board (IRB) before the study commences. This includes both primary data

collection and secondary data retrieval. When submitting to IRB, researchers must provide proof of human subjects training and all documents (e.g., surveys, interview questions, recruitment material) related to the study. Approval by the IRB helps ensure all research is conducted in accordance with federal, state, and local regulations and guidelines which protect human participants, students, and staff involved with research. Research studies have a one-year IRB approval timeframe. Should the study not be complete within one year, researchers submit a request for continuation each year until the study's completion. Two of the current 10 members of the university's IRB are School of Education faculty members.

- Whether your research outcomes have ever been challenged based on the validity of the methods
  It is fairly common for others to attempt to replicate or prove/disprove results based on their sample; this is a part of research. However, to our knowledge, studies have not been challenged based on the validity of the methods. Our school is a very collaborative one, and researchers often discuss their studies with others, raising any concerns or uncertainties so feedback can be provided, before conducting the studies. This helps strengthen the research design and methods and helps safeguard the study from being challenged based on validity concerns.

RESEARCH QUALITY AND INDEPENDENCE AT MATHEMATICA

Quality assurance process

Mathematica has well-defined procedures for conducting effective quality assurance (QA) reviews of all research methods and products. To develop high quality products that meet the needs of the project, our QA reviews focus on responsiveness to the audience, the correctness of the methods used, the appropriateness of the interpretation and conclusions of the results, completeness, and clarity. Mathematica has developed specific QA guidelines for a range of products and research activities that are likely to be part of Research Practitioner Partnership (RPP) project activities, including for data collection plans, high-stakes systems, program technical assistance, and data processing and programming. Although some details might differ based on the needs of the project, our five-step approach to QA will generally apply to all tasks and deliverables:

1. **Planning.** For each RPP research project that Mathematica undertakes, the Mathematica lead will draw on our established QA guidelines relevant to the project's needs to develop a QA plan to guide review of planned research methods and deliverables. The QA plan will reflect our understanding of the work, potential risks, the quality control steps to implement to reduce risk and ensure quality. The plan will also reflect the specific needs of the project.
2. **Implementation.** The Mathematica lead will review each deliverable to ensure it aligns with the goals of the research project and is consistent with the goals and approach of the RPP. We will not submit draft or final deliverables unless both the Mathematica lead and the appropriate QA reviewer (Step 4) have a high degree of confidence that the findings have passed rigorous and robust QA and quality control procedures.
3. **Monitoring and iterative updates.** The Mathematica lead will revise the QA plan at least annually, based on any changes to the work or new potential risks.
4. **Comprehensive QA and submission.** As a deliverable nears completion, Mathematica will conduct a comprehensive QA review by experts who are not involved in day-to-day project work. This independence provides an objectivity that enhances the quality of Mathematica's work. The reviews focus on whether the deliverables are of high quality and meet the needs of the project. The project team will then implement the reviewers' feedback and submit formal deliverables for editorial review and production.
5. **Lessons learned.** Throughout the QA process, the Mathematica lead will gather input from staff about how the QA plan worked or could be improved for the next iteration or cycle of the task. At the conclusion of tasks and/or deliverables, the lead will document findings regarding process improvements and risk monitoring so that they can lead to improvements on other Mathematica-led RPP research projects.

Identifying and addressing data quality limitations

Mathematica has documented procedures that researchers and programmers implement to assess the quality and completeness of data that we will use to complete project activities and address any limitations in the data. These procedures include (1) developing for each research project a set of data requirements, including the data elements and samples, that are necessary for completing the planned analyses; (2) sharing a formal data request with the data provider that outlines the data requirements; (3) discussing the data requirements with the data provider to identify any known

limitations or concerns and address the issues by making any necessary accommodations in the analysis plan and data requirements; (4) inspecting the data and developing and implementing a series of steps to check that the data correspond with the data requirements; and (5) discussing any questions about the data that arise while implementing these steps with the data provider, documenting resolutions to these questions, and revising the analysis plan, if needed.

Accountability, transparency, and independence in our work

Uncompromising objectivity and quality are core values that guide our work. We monitor the objectivity and independence of our work through our QA process, in which experts who are not involved in day-to-day project work comprehensively review all research products. We will also submit all research products for external peer review by the RPP and carefully address all concerns identified in the review. In addition, our products provide detailed descriptions of our research methods that are transparent and complete.

When there is an additional risk or perception that the independence of our work might be compromised, we will take the following steps, as appropriate. First, when developing a plan for conducting the research, we will consider partitioning the project team. For example, in our study of the Teacher Incentive Fund, we used distinct teams to conduct work to better implement and to evaluate the program. We will also enhance our QA of the research methods and products by engaging another expert from within Mathematica who has no previous relationship with our work in DC or other relevant clients to conduct a review focused on identifying and addressing risks related to the product's independence and objectivity. Finally, we will take additional steps to demonstrate transparency in our work, which could include preregistering an analysis plan—as we did for our study of postsecondary impacts of KIPP middle schools—or sharing public use data files and code used to produce the results.

Obtaining institutional review board clearances

Mathematica is experienced in obtaining clearances from institutional review boards for the protection of human research subjects. Our staff will determine whether a project requires institutional review board clearance and prepare necessary materials. Mathematica is accustomed to working with multiple institutional review boards and to adapting materials to meet varied requirements. We have established relationships with some boards and will prepare materials to obtain clearance from state or local boards when needed. Our staff will review and prepare supporting materials, ensure that we meet submission schedules, assist with and respond to institutional review boards' questions and stipulations, and track approval status.

Challenges to the validity of research methods and results

Mathematica is firmly committed to procedures that support the accuracy and integrity of its products and results, investing heavily in ongoing QA and testing designed to minimize the chance of error. Mathematica fully and transparently investigates any challenges to the accuracy of a product and proposes an appropriate resolution whether an external party or Mathematica staff identify the concern in the course of QA efforts. For example, in December 2013, Mathematica staff performing ongoing quality review discovered a coding error in programming code used to develop value-added estimates for DCPS teachers. We immediately notified DCPS of the error, revised the programming code, reran the model with the revised code, and provided the new teacher value-added scores to DCPS so it could determine how this affected its teachers. The revised programming code resulted in small changes in the value-added scores of teachers, resulting in changes in the performance

ratings of 44 teachers (out of 4,400); 22 teachers had their ratings adjusted upward and 22 had ratings adjusted downward. Shortly thereafter DCPS alerted affected teachers and their principals. Mathematica is committed to producing high quality work and we deeply regret the coding error. To prevent similar errors in the future, we strengthened our review and testing in projects with complicated data cleaning and processing tasks, particularly those in which our work is associated with high stakes for individuals or organizations. We developed supplemental QA guidelines for these high-stakes projects, which included implementing phased gate reviews that evaluate progress at defined steps, including staff from outside of the project team in programming code reviews, and parallel coding by independent programmers for the most sensitive efforts.

Trinity

Research Quality and Independence

Research at Trinity is conducted by faculty members in their fields of expertise, interest, and training. All faculty, including those in the Education Program, have doctoral degrees with extensive training in research methodology and data interpretation; this helps to ensure that any research conducted by the members of the Trinity academic community is valid and has high-quality methods and outcomes.

All research projects are submitted to the Institutional Review Board, which reviews them for adherence to ethical standards and also checks that the principle investigator has the appropriate qualifications and experience for the project, evaluates the ongoing evaluation of the project, and assures that there is a plan for the recording and maintenance of all data. In addition to these internal protocols, as an academic institution faculty are engaged in research that is peer-reviewed by colleagues in their disciplines. Faculty share their research findings at professional conferences and publish in peer-reviewed journals and therefore the processes inherent in publication process further check validity and quality of methods and outcomes.

Trinity avoids variations in internal data by ensuring that data is compiled in and pulled from multiple databases, which allows results to be triangulated. Senior Staff members review data before drawing conclusions or sharing findings and if there is any reason to lack confidence in the data quality then it is not used.

Members of the Academic Affairs team and other Senior Staff members review any final deliverable products. In addition, as noted – faculty research products are peer reviewed as part of the publication process.

Trinity's compliance and risk management program is central to its academic mission and to the administrative structure that supports it. Transparency and ethical conduct are expected of all members of the Trinity community, and are embodied in our policies on fundraising and grant administration, "whistleblower" procedures, policies on the administration of contracts, prohibition on self-dealing or any potential conflicts of interest, and policies that direct the observance of all federal regulations surrounding financial transactions. All of these policies are published on Trinity's website, and are reviewed frequently with senior staff and refreshed as necessary.

As an institute of higher education, Trinity has an Academic Freedom policy in line with the 1940 Statement on Academic Freedom of the American Association of University Professors and the Association of American Colleges. In regards to research, this policy states that “full freedom in research and in the publication of the results” which ensures complete independence. All grants, including those with a research component and/or outcome analysis, is reviewed by members of the Academic Affairs team and Trinity’s President to ensure that complete independence is maintained.

Trinity has a full Institutional Review Board and associated policy statement. This policy statement and the IRB process applies to research projects conducted by faculty or students of Trinity that involves human subjects and applied to research that faculty and students of Trinity

conduct on its own premises as well as with or through other institutions or any external locations. The IRB policy specifies three major principles, which are taken from the Office for Human Research Protection and undergird its guidelines: respect for persons, beneficence, and justice. An IRB board, consisting of five members who are sufficiently qualified though their experience and expertise. This includes at least one member who is not otherwise affiliated with the institution. Members are appointed to serve on the IRB by Trinity's President in consult with the Provost's office. All research projects conducted by members of the Trinity community that are not otherwise deemed exempt are required to submit an IRB application which the IRB Board then review according to the guidelines for IRB decisions stated in Trinity's policy.

UNIVERSITY of VIRGINIA | CURRY SCHOOL and BATTEN SCHOOL EdPolicyWorks

EdPolicyWorks - Research Quality and Independence

Data quality is the foundation on which research is based. EPW researchers take great care in the acquisition, management and use of data. As we acquire data we typically build a data dictionary that documents the definition of each variable, its source and the key moments of its distribution, e.g., means, standard deviations, minimums and maximums. We then run both internal and external validity checks on each variable to assess its accuracy, e.g., consistency across variables within a database and between databases. Typically, we have multiple researchers examine the code and outcomes to minimize mistakes. Given our substantial experience across multiple sources and databases, as well as the same sources and databases over time, we have developed coding and internal validity checks that minimize errors. As we employ data for analysis, we assess whether the outcomes produced are consistent with outcomes from other, similar, research. When they are not, we examine whether the differences make sense in light of theoretical or contextual differences.

Some data are less reliable than other data. When we employ data that are judged to be less reliable, we go back to the source to find other ways to triangulate better measures. Depending on our confidence in the resulting data, we employ it more skeptically to assess whether potential inaccuracies influence our outcomes.

Our research products are reviewed by senior researchers with substantial experience employing data across a variety of situations. In nearly all cases our analyses are subject to rigorous peer-review by other researchers as we submit the research to peer-reviewed journals.

EPW researchers maintain an arm's length independence from our policy collaborators. We do not see ourselves as consultants but rather as collaborators. This understanding is typically made clear in a data use agreement that stipulates that EPW researchers retain editorial control over all research products. To maintain both the perception and reality of independence, we rarely accept funding from the organizations with whom we collaborate. We typically receive financial support from the federal government or foundations who support our research based on its merits. To ensure accountability and transparency we make public nearly all of our research and most of this publicly available research is published in peer-reviewed journals.

All of our research is subject to the norms of ethical treatments of subjects and research methods as codified by the University of Virginia Social and Behavioral Institutional Review Board, which must approve each of our projects.

Our research outcomes have never been challenged based on the validity of the methods employed in the research.