Building an Early Childhood Database for the Global Research Community

The New Jersey Families Study

Thomas Espenshade

Office of Population Research

Princeton University

PRINCETON UNIVERSITY

The Setting

- Persistent gaps in opportunity affect access to higher education, the quality of the U.S. workforce, and patterns of adult social and economic inequality.

- More than half of low-income children in America enter kindergarten lacking the math, verbal, and social-behavioral skills needed to succeed.

- Black children are on average already one year behind white children when they begin kindergarten. The gap is four years by the time of high school graduation.

- These pre-K achievement gaps are of national concern and have attracted the attention of policy makers.

Reform Efforts

- Most reform efforts occur too late.
- Federal programs such as *No Child Left Behind* and *Race to the Top* focused on K-12 school reform.
- The emphasis on charter schools, smaller class size, extended day, year-round schools, higher pay, and more effective teachers is also focused on K-12, not pre-K.
- The push for universal pre-K recognized that reform efforts need to begin before kindergarten.
- The Head Start program enrolls low-income, pre-school children in an effort to boost school readiness. Early Head Start offers similar services to infants and toddlers.
- Nevertheless, despite billions of dollars expended by American taxpayers, corporate philanthropy, and the nonprofit sector to achieve greater parity, income gaps in student outcomes have risen while racial and ethnic gaps have declined only slowly.

The Emerging Focus on 0-3

- Greater attention is now being directed to children's youngest ages – zero to three – when the brain has the greatest capacity for growth and development.

- What is it about these earliest years that is so critical? There is growing consensus around the importance of family dynamics and the home environment in helping children get off to a strong start.

- "If there's one overarching lesson from the past few decades of research about how to break the cycles of poverty in the United States, it's the power of parenting – and of intervening early, ideally in the first year or two of life." N. Kristof and S. WuDunn, *The New York Times*, 9/12/2014.

Families as Schools

- Families are small schools, and parents are their children's first teachers.
- Every child in America is being home schooled in the sense that children's expectations, aspirations, and early abilities are shaped at home.
- Children's behaviors, learned skills, and knowledge are forged in the crucible of parent-child interactions.
- Yet we know remarkably little about the nature, frequency, or quality of these interactions.
- These lessons – the teaching and learning that goes on at home – are typically hidden behind closed doors.

The Vision

- Imagine a world where we had a window into the daily lives of American families and a rich understanding of how families support their children's early learning.

- Imagine a world where the diverse strategies that families from a wide variety of backgrounds use to build skills in their pre-school children are revealed to us in everyday parent-child interactions.

- Imagine a world where children could be lifted from a path of cumulative disadvantage onto a path of cumulative advantage.

- In such a world, the legacy of early disadvantage would not necessarily follow children into adulthood and reproduce itself from one generation to the next.

- Such a world is possible with the New Jersey Families Study.


New Jersey Families Study
www.njfamilies.princeton.edu

The New Jersey Families Study

- We endeavor to learn how families support their children's early learning.
- Our study features a highly innovative "video ethnography" that captures families' daily activities at home.
- High-definition video cameras with microphones are located strategically in up to four rooms in participants' homes – rooms where most parent-child interactions occur.
- Up to eight cameras are activated simultaneously and continuously throughout the day and evening for two weeks.
- No researcher is present while video recording takes place.

An Intense Focus on Parent-Child Interactions

Interactions that hold particular interest are those believed to be linked to cognitive and social-emotional development:

- The amount of talking and reading parents do with children.
- Children's sleep patterns and practices; their diets and nutrition.
- Their exposure to electronic screen time.
- Structure and routines at home.
- The way that stress outside the home affects parenting practices.

The NJFS Data

- The New Jersey Families Study (NJFS) video recorded the in-home activities and behaviors of 21 families in Mercer County, New Jersey over a two-week period.

- The sample is typical of young families with two- or three-year-old children and highly diverse by race and ethnicity, social class, family structure, and place of residence.

- The data capture more than 5,700 hours of the daily lives of families – approximately 275 hours per family – with particular attention to parent-child interactions. (Currently, the video data reside in nearly 463,000 discrete video clips, each with date and time stamps, because the cameras were on motion sensors to conserve recording space.)

- The project's video data are supplemented with survey and interview data collected during six additional points of contact with families.

The Virtues of Video Data

- Video recording reduces the social desirability bias that can sometimes surface in respondents' answers. Researchers are freed from reliance solely on survey or interview data filtered through parents' reports of their "Instagram" selves.
- Participants need not remember what happened or when.
- Using unobtrusive technologies is likely to reduce interviewer effects and produce more reliable measures than surveys or other simpler in-home methods.
- A two-week window of observation lowers the likelihood that participants will be on their best behavior all the time.
- Viewing families in their daily routines has the potential of serendipitously capturing events and behaviors that investigators might not have thought to ask about otherwise.
- The ability to replay and recode video data opens the data to a wider set of disciplinary and cultural interpretations.
- Comparing what parents say they do with what they actually do can shed new light on the reliability of participants' answers to conventional surveys and interviews.

Contributions to Early Childhood and Children's Later Years

- Our study breaks new ground. It is the first time anyone has attempted an in-home naturalistic observation of this breadth, intensity, or duration.
- This holistic approach to parent-child interactions and a deeper understanding of the contextualized strategies that families of different backgrounds use to support their children's early learning will help more children become school-ready.
- It will also lead to a better appreciation of the daily struggles facing many families and thereby help schools become more "children-ready."
- It will help to narrow the widening gap between high and low achievers by illuminating children's most formative years when those gaps begin.
- Demonstrated success for the New Jersey Families Study will encourage researchers to replicate this innovative project in other communities, helping to create a more nationally representative early childhood database.

Broader Significance

This study presents significant, groundbreaking opportunities:

- Opening up the data to a global community of scholars maximizes the return on investment. Princeton University has limited resources in the area of early childhood, whether in terms of faculty or professional schools. Taking all of these video clips and auxiliary survey and interview data and constructing a database that can be used and analyzed by researchers around the world enhances substantially the value and potential of the data.

- This truly unique study joins forward-looking data with a host of cutting-edge data science techniques. The latest advances in computer vision, speech recognition, artificial intelligence, and machine learning would allow us to handle a staggering amount of data, to create a user-friendly database, and to roll it out in a timely fashion.

- This study could serve as a model for future research projects that aim to open up huge amounts of digital data (video, audio, etc.) to a larger research community.

- Funding could lay the groundwork for the development of totally new types of secure data-sharing platforms.

- Funding for the project would enable these data to be housed in a robust environment that would facilitate future research using methods that have yet to be developed.

The Request

To build a New Jersey Families Study Early Childhood database for a worldwide research community, as a next step we seek a two-year planning grant of $300,000 to:

- convene a series of workshops and focus groups with potential users of NJFS data to obtain their feedback on desirable features of end products.

- conduct a variety of tests and experiments with a sample of the data. Tests would focus on tagging and stitching discrete video clips and transcribing audio tracks. These experiments would allow us to determine the optimal approaches, including time and money costs, for preparing a sustainable database that houses all related NJFS data and documentation.

- continue working with colleagues in Princeton's Research Computing program to develop software and security procedures to enable researchers outside Princeton to have remote access to NJFS data without being able to copy them.

- prepare a detailed proposal to fund the application of what we learn during the planning period to all of the NJFS data and make them available on a restricted-access basis to the global research community.

Princeton's NJFS Collaborative Network

Thomas J. Espenshade, Ph.D., Senior Scholar, Office of Population Research and NJFS Project Director

Tom Espenshade led the NJFS data collection effort and is now assembling a team to make the data available to a global research community. He is the author (with Alexandria Radford) of *No Longer Separate, Not Yet Equal: Race and Class in Elite College Admission and Campus Life*.

Drew Allen, Ph.D., Executive Director, Center for Data Driven Social Sciences

Allen is leading the day-to-day efforts to enhance the support of data computation and information technology intensive social science resources at Princeton. One initiative is piloting a secure data analysis virtual environment.

The Center for Digital Humanities is an interdisciplinary research center and academic unit within the Princeton University Library.

They embrace an inclusive understanding of Digital Humanities that investigates the myriad ways digital methods and technologies are opening avenues for research into the human experience, past and present.

Princeton's NJFS Collaborative Network

Center for Statistics and Machine Learning

The center collaborates on research that combines insights from computation, machine learning, and statistics with specific application domains. It also supports innovations in the theoretic foundations of data science, including advanced algorithms for big-data problems, machine learning, optimization, and statistics.

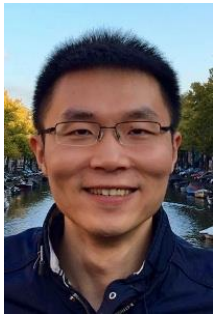
Wind Cowles, Ph.D., Director, Princeton Research Data Service

The PRDS initiative was launched in May 2019 to provide the Princeton research community with the expert services and infrastructure needed to store, manage, retain, and curate digital research data, and to make digital research data available to the broader research community and to the public. Her research interests include the influence of sociolinguistic factors on speech production in children and adults.

Data and Statistical Services (DSS) provides data and statistical consulting.

The service is located in Firestone Library.

Princeton's NJFS Collaborative Network



Jia Deng, Ph.D., Assistant Professor of Computer Science and Director, Princeton Vision and Learning Lab

His research focus is on computer vision and machine learning, in particular, achieving human-level visual understanding by integrating perception, cognition, and learning.

Kathryn Edin, Ph.D., Professor of Sociology and Public Affairs, Woodrow Wilson School, Co-director of the Bendheim-Thoman Center for Research of Child Wellbeing (CRCW)

Edin is co-PI of the American Voices Project. The project will interview 5,000 individuals across the country to understand their successes and struggles, their hopes and dreams, and their ideas for making our country work better.

Ed Freeland, Ph.D., Associate Director, Survey Research Center (SRC); Lecturer in Public and International Affairs

Dr. Freeland teaches a graduate seminar on survey research methods and advises faculty and students who are conducting survey research projects. He is a former member and Chair of Princeton's IRB panel.

Princeton's NJFS Collaborative Network

Curtis W. Hillegas, Ph.D., Associate CIO, Research Computing, Office of Information Technology

Hillegas has assisted Princeton's Fragile Families Project, which is restricted-access but all text data. He is eager to become involved with the NJFS because of the new challenges that working with audio and video data bring.

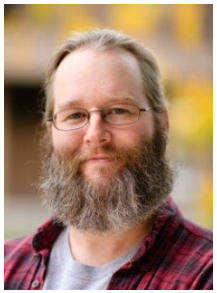
Jeffrey Himpele, Ph.D., Director, VizE Lab for Ethnographic Data Visualization; Lecturer in Anthropology

His interests include visual anthropology and documentary filmmaking; media, sound and data visualization; capitalism and commodification; indigenous media and politics; Bolivia and the Andes.

David Hopkins, Senior Manager, Video Production Support, Office of Information Technology

David and his team produce video projects that range from single camera interviews to multi-camera live productions. His team was responsible for installing the video recording equipment in participants' homes for the NJFS data collection.

Princeton's NJFS Collaborative Network



Steve Huston, Infrastructure Operations Analyst/Manager, Astrophysical Sciences

Steve works alongside Curt Hillegas to support Princeton's Research Computing.

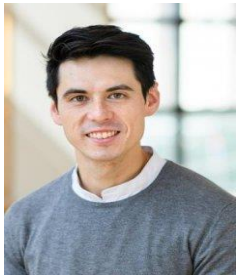
Jennifer Jennings, Ph.D., Professor of Sociology and Public Affairs, Woodrow Wilson School, Director of the Education Research Section (ERS)

Her research interests include educational and health inequality; race, gender, and class disparities; and organizational accountability systems.

Ben Johnston, Senior Educational Technologist, Education and Classroom Technologies, McGraw Center for Teaching and Learning

He works with faculty and students across the University to facilitate the integration of digital platforms and implementation of technology tools into coursework.

Princeton's NJFS Collaborative Network



Dean Knox, Ph.D., Assistant Professor in Politics

He developed quantitative models and methods for new forms of social science data, including path data for sequential decision-making and audiovisual data conveying human emotion. Dean has extensive experience with machine transcription of audio data.

John B. Londregan, Ph.D., Professor of Politics and International Affairs

Londregan is a specialist in the development and application of statistical methods in political science, especially text as data.

Nolan McCarty, Ph.D., Susan Dod Brown Professor of Politics and Public Affairs

Nolan McCarty is the faculty director of the new Data-Driven Social Science initiative. Nolan will lead and guide our efforts to test out ideas and explore strategies for encouraging and supporting the use of data (including "big data") in academic research in the social sciences at Princeton.

Princeton's NJFS Collaborative Network

Karthik Narasimhan, Ph.D., Assistant Professor of Computer Science

His research spans the areas of natural language processing and reinforcement learning. He is especially interested in developing autonomous systems that can acquire language understanding through interaction with their environment.

Steve Niedzwiecki, Associate Chief Information Security Officer, Information Security Office, Office of Information Technology

Seth Porter, Master of Library and Information Science, Head of Donald E. Stokes Library

Seth is the driving force behind the Data Visualization Lab. The lab provides NVivo and other qualitative and quantitative software and supplies graduate student help on various software platforms.

Princeton's NJFS Collaborative Network

Boriana Pratt, Statistical Programmer, Office of Population Research

Boriana is the data manager for the New Jersey Families Study, responsible for receiving and securely storing project data. She also serves as the liaison to the Broadcast Center within Princeton's Office of Information Technology.

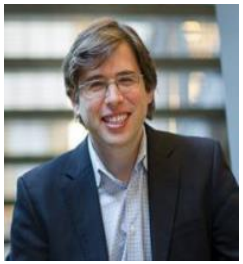
Peter J. Ramadge, Ph.D., Gordon Wu Professor of Electrical Engineering and Director of the Center for Statistics and Machine Learning

His research interests are in the areas of signal processing and machine learning. He works on a variety of fundamental problems (including boosting, adaptive signal processing, and learning from data) and in a variety of application domains (including fMRI analysis and video analysis, annotation, and search).

Olga Russakovsky, Ph.D., Assistant Professor, Department of Computer Science

Her primary research area is computer vision, closely integrated with machine learning; human-computer interaction and fairness; accountability and transparency.

Princeton's NJFS Collaborative Network



Matthew Salganik, Ph.D., Professor of Sociology and Interim Director, Center for Information Technology Policy

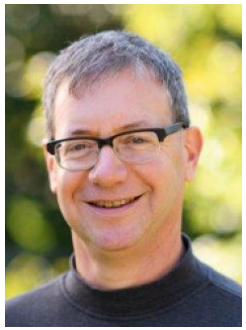
His research interests include social networks and computational social science. He is the author of *Bit by Bit: Social Research in the Digital Age*.

David Sherry, Chief Information Security Officer, Office of Information Technology

Brandon Stewart, Ph.D., Assistant Professor, Department of Sociology

His research interests include computational social science; computer-assisted text analysis; Bayesian statistics. He is also affiliated with the Politics Department, the Office of Population Research, the Princeton Institute for Computational Science and Engineering, and the Center for the Digital Humanities.

Princeton’s NJFS Collaborative Network



William Wichser, Associate Director, Systems and Storage, Office of Information Technology

Nick Budak, Web Developer, Center for Digital Humanities

He enjoys imagining and implementing accessible, dynamic interfaces for digital humanities projects. Nick is a Certified Professional of Web Accessibility through the IAAP.