

NORTH CAROLINA EDUCATION RESEARCH DATA CENTER
Grant # 200300138
Final Report to the Spencer Foundation

Clara Muschkin, PhD
Kara Bonneau, MA
Kenneth Dodge, PhD

Center for Child and Family Policy
Duke University

June 15, 2011

NORTH CAROLINA EDUCATION RESEARCH DATA CENTER

Final Report 2009-2011

Executive Summary

In 2009, the Spencer Foundation renewed its generous support of the Data Center with an additional two years of funding. In addition, Duke University and the North Carolina Department of Public Instruction have continued to support the Data Center through the collaborative relationship established in the current Memorandum of Understanding between the two institutions.

The years covered by this Spencer Foundation grant, 7/2009 through 6/2011, were very productive for the Data Center. During this the Data Center received and processed over 50 new data sets with data on over 1.3 million students, 102,000 teachers, and 2,400 schools in North Carolina. These files were cleaned and integrated with existing data files from previous years. They have been made available to both junior and experienced scholars for original research.

During this period, the Data Center received 214 new research inquiries from faculty members and students (144 since March 2010). A wide variety of topics is being investigated, ranging from the impact on student performance of investment in early childhood programs, to identifying the effects of state accountability sanctions on specific disadvantaged student sub-groups, to exploring the success of women and minority college students in STEM majors.

One of the priorities during the past grant period has emphasized outreach to academic institutions in North Carolina that are collaborating with state agencies to evaluate key programs. Another priority is the linking of education data to individual-level data from other administrative sources. We have continued to work on expanding our data holdings, especially the acquisition and integration of data files on student outcomes after high school. During 2009-11, we acquired or updated and matched school data with individual records on community college experience, as well as on juvenile crime, arrest, and incarceration. During the current year, we also have successfully acquired and matched data from detailed birth records to school records, for births occurring in North Carolina since 1987.

History and Function

In 2000, an interdisciplinary group of researchers from Duke University and the University of North Carolina received financial support from the Spencer Foundation to establish the North Carolina Education Research Data Center (Data Center or NCERDC). At that time, the North Carolina Department of Public Instruction (DPI) joined us as a partner in creating and operating the Data Center. The Data Center's goal is to support empirical research on topics relevant to education policy by making available data sets on students, teachers, and schools.

In 2003, the Spencer Foundation renewed its funding for this project with a four-year award. In November of 2008, Duke University and DPI renewed the Memorandum of Understanding that permits this arrangement for data sharing, for at least another three years. The Data Center underwent a change in leadership in July, 2008, when Clara Muschkin, Assistant Research Professor of Public Policy, was named NCERDC Director and Kara Bonneau was appointed NCERDC Associate Director. In July 2009, the Spencer Foundation renewed its support of the NC Education Data Center with a two-year grant, with Kenneth Dodge as Principal Investigator and Clara Muschkin as Co-Principal Investigator. The project is entitled "Longitudinal Data for Education Reform: A Critical Role for the NCERDC."

The Data Center supports a vast array of academic and policy-oriented research projects. Since January 2001, the Data Center has provided data for 103 faculty research projects, 33 of these undertaken by faculty or senior researchers from institutions outside of North Carolina. The past two years were particularly active for the Data Center, with an increase of 38 new faculty projects. Since 2001, the Data Center also has supported 70 graduate student projects, primarily for doctoral dissertation research. Of this grand total, 22 new student projects were approved during the past grant period.

During the past two years, the Data Center received 144 new research inquiries from faculty and students from across the United States. Scholars have reported results in professional conferences, academic journals, policy briefs and reports. Topics of study have included the minority achievement gap, attracting and retaining qualified teachers, school accountability, effectiveness of charter schools, the academic performance of at-risk children, problem behavior in schools, the impact of new middle and high school programs on student achievement, the effectiveness of programs for gifted as well as special needs students, and life-course outcomes of educational experiences.

Having acquired thousands of files, Data Center staff members have encrypted confidential information, created longitudinal databases, and linked district, school, classroom, student, and teacher data files. Converting, formatting, linking, and documenting these large, complex files require substantial investments of time. The Data Center staff completes this work, and stores and maintains these files in a centralized and secure location. The Data Center maintains a website with detailed descriptions of the data files, eligibility requirements for researchers who wish to obtain access to these data for their research, and data use application procedures. The NCERDC Director and staff members provide a great deal of guidance to scholars in designing their projects and considerable mentoring of graduate students who use these data for their dissertation research. They also engage in outreach, providing information to

academic researchers about availability of the data and the value of longitudinal information for policy analysis. Often, scholars are limited in the research questions that they ask by their assumptions about the data that are available for analysis. By communicating to scholars about the vast data files that could be made available to them, and about the cutting-edge research that employs these data, we improve the quality of research questions that scholars are able to ask.

Data Acquired

Working with the North Carolina DPI, the Data Center has acquired various files pertaining to districts, schools, students, and teachers. DPI annually collects data on its 115 districts, 2,400 schools, 1.3 million students, and 101,000 teachers. Every year, the Data Center acquires the most recent school year's records and matches teacher and student records to those from prior years so that researchers can track teachers' careers and students' academic progress. During the past year, data center staff created and updated more than 100 datasets. The Data Center has procedures to identify a student's peer group, geocode the student's address, aggregate student or teacher data to the school or district level of analysis, and link school, teacher, and student data to files from other sources. Having detailed longitudinal information about students, teachers, and schools provides an excellent opportunity to examine the outcomes of education policies.

During the past two years, the Data Center integrated several new sets of new data files, including:

- **Absences and Demographics** contains information on student demographic characteristics and absence from school. This file provides the basis for membership in grades 3 through 13. These files contain basic student demographics as well as detailed information on membership and days absent.
- **New course membership** listings provide detailed student-level information on course-taking, including course title and instructor's name. These new files include information on students across all grades from PK through 12, whereas prior files of this type did not include students in grades PK through 2. These data will be very useful to researchers who are exploring such questions as: student mobility, preschool attendance, and grade promotion/retention in the early grades.
- **ASSET Test Score** data derive from a series of writing, reading and math test used by community colleges for admission and placement and used to measure the academic attainment of career and technical education students as required by federal regulations.
- **GPA Student Data** information on 12th grade student GPA, class rankings, and high school completion, with data for nearly all students in the 2008-09 and 2009-10 academic years.

- **High School Transcript Data** include information from students' high school transcripts, including courses taken, credits, class ranking, and course grades. Information is available only for a small number of districts in 2005, and increases over time; with data for nearly all students in the 2008-09 and 2009-10 academic years.

Appendix A provides a description of all datasets in the NCERDC archive.

The Data Center has continued to work toward expanding our holdings, to include information from post-secondary education institutions in North Carolina. We have moved forward with efforts to link the DPI elementary and secondary school data with records from the North Carolina Community College System. Such matches allow researchers to determine who follow could follow high school graduates through their college achievement and attainment, and will provide education policymakers with important information on enrollment and success of students in the state's largest post-secondary institution. In conjunction with the leadership of the NCCCS, we have developed procedures to link and share these datasets for ongoing research projects. The linked data files are being used to examine questions that are of high priority for the leadership of the NCCCS, such as predictors of remediation and placement assessments upon entry to community college.

The Data Center also is making progress toward expanding its holdings through integration of data from other state agencies. We have worked with the North Carolina Department of Vital Statistics to access detailed birth records from 1987 through 2009, and merged them to the student records in Data Center files. Both children and young parents can be identified in both files and their records matched, in order to examine questions regarding the long-term impact of birth outcomes, predictors of teen pregnancy, and impact of parenthood on education outcomes.

The Data Center has successfully merged files from the North Carolina administrative data with other files to create innovative approaches for studies. For example, in the evaluation of the Early College High Schools, the Data Center is linking application data and subsequent opinion surveys to the administrative education data. In an evaluation of the North Carolina Child and Family Support team program, the Data Center is tracking the academic history and progress of students served by the program, identifying control groups, and linking student education data with information from state agencies such as Department of Social Services and Juvenile Justice. For a study of the relationship between participation in preschool for disadvantaged four year olds and subsequent academic performance, the Data Center has linked student records from the More at Four Evaluation Study with the administrative education data. For the numerous studies examining outcomes beyond schooling, the Data Center has linked student information with records from the NC Department of Correction, Office of Juvenile Justice and Delinquency Prevention, Vital Records, and the NC Community College System.

New Research Projects Using these Files

The Data Center continues to provide ongoing support to multi-year research projects. The Data Center's contribution to the evaluation of the state-funded Child and Family Teams

project will continue at least through 2010. The NC Education Research Data Center is a key component of two multi-year awards from the U.S. Department of Education, and is a key component of a new award to Duke researchers from the Smith Richardson Foundation. The NCERDC leadership are actively involved in each of these projects, and the staff work closely with the associated researchers to meet their data needs.

- CALDER. The Center for Analysis of Longitudinal Data in Education Research. This is a federally-funded national education research center with scholars from the Urban Institute, Duke University, Stanford University, the University of Texas-Dallas, the University of Florida, the University of Washington, and the University of Missouri. CALDER studies involve cross-state collaborations among researchers on studies of the outcomes of teacher policies, governance policies, and social and economic community conditions.
- Early College High School Evaluation. This is an experimental study of a complex high school model that is being widely implemented throughout the United States. Through the use of random assignment of students to schools, the research project will be able to make causal inferences about the impact of the Early College Model on student achievement and attainment. We have worked with this research team to develop a proposal for additional grant funding.
- Schooling and Life-Course Outcomes in Early Adulthood. The project examines the connection between test scores and important outcomes that signal or influence the accumulation of human capital. The research will have policy implications for high school graduation and post-secondary education rates, labor market participation, early child bearing, and delinquency and criminal activity patterns. The NCERDC is a central component of this project.
- NIDA Transdisciplinary Prevention Research Center. The mission of this center is to support research that develops evidence-based prevention strategies addressing problematic substance abuse in teenagers and young adults. The NCERDC is involved in many of the projects supported by this center.
- The Children's Environmental Health Initiative explores environmental justice dimensions of childhood exposure to hazards such as lead and air toxins. The NCERDC is linking student-level records to pediatric data, birth record data, and blood lead surveillance data for the researchers to examine the relationship between educational outcomes and spatially identified exposure risks.

Additionally, the Data Center has supported numerous other innovative studies from multiple universities and departments, including the following:

- Maria Fitzpatrick is examining the impact of three teacher incentive programs implemented in North Carolina on long-term student achievement.

- Douglas Lauen is estimating the effects of subgroup specific accountability for academic gains among disadvantaged students.
- Kirabo Jackson is investigating the mechanisms through which North Carolina charter schools recruit highly qualified teachers.
- Alan Maloney is evaluating the effectiveness of a professional development model for mathematics education in NC STEM schools.
- Beth Gifford is leading an ongoing project to evaluate the effects on students and teachers of a school-based intervention program based on child and family support teams.
- Marie Lynn Miranda is examining the predictors and consequences of family mobility during early childhood for subsequent health and educational outcomes.
- Jane Hannaway is the NC data to identify the policies, programs, and practices of turnaround schools that have achieved rapid improvements in student outcomes.
- Martha Putallaz is investigating the links between preadolescent girls' friendship and peer relationships and their academic and social outcomes in school.
- Charles Clotfelter and colleagues are using a data file that links community college student records to those of the NCERDC, to study enrollment and success in community college.
- Clara Muschkin and Kara Bonneau are examining the long-term impact of participating in a high school transition program for multiple high school outcomes.
- Katy Rouse seeks to evaluate the impact of year-round schooling on at-risk students' academic outcomes.
- Helen Ladd and Charles Clotfelter are examining the student-level implications of school and residential segregation in the two largest NC counties.
- William Darity and Kara Bonneau are studying patterns and determinants of advanced course taking among NC high school students.
- Elizabeth Stearns is examining the factors that contribute to the success of women and minorities in math and science majors in NC universities.
- David Ribar is evaluating the changes in student attendance and academic performance associated with policy changes in the provision of school breakfasts through the US Department of Agriculture School Breakfast Program.
- Helen Ladd, Kenneth Dodge, and Clara Muschkin are studying the impact of birth outcomes and exposure to early childhood programs in NC counties on educational outcomes for individual students.

Appendix B has a bibliography of all faculty projects and abstracts for the newest projects.

Student Research Projects

This past year, the Data Center has supported several new student research projects focusing on issues including the impact of school practices, attracting and retaining qualified teachers and principals, and consequences of No Child Left Behind. Some examples include:

- Mitchell Hoffman is examining the educational consequences of recent growth in rates of incarceration.
- Danielle Li is studying the allocative efficiency of principal and teacher matches to schools, and how wage contracts can be designed to enhance their fit.
- Erika Martinez is exploring the response of housing markets to performance-based and growth-based measures of school quality.
- Solomon Pantelis is studying the consequences of a policy change to include dropout rates as an assessment criterion for high school accountability.
- Ken Whelan is examining the effects of different grading standards across high school courses for student achievement.
- Ben Ost is focusing his research on curriculum familiarity as a mechanism through which teacher experience may contribute to student success in the classroom.
- Nirav Mehta is developing a model of competition between public and charter schools as a framework for estimating the long-term effects of charter school entry on student test scores.
- Patten Mahler is identifying the influence of pension-related incentives for teacher retention and retirement decisions.
- Sara Pilzer is focusing her research on the causes and consequences of policy changes related to school attendance boundaries, for students who are re-assigned as well as for the students who remain in their original schools.
- Kevin Schnepel is evaluating the potential biases in education research associated with the use of fixed-effects regression in the presence of heterogeneous treatment effects.
- Richard Mansfield examines the implications of teacher sorting across and within high schools for inequities in student academic performance.

Appendix C has a bibliography of all student projects, with abstracts for the newest projects.

Examples of Recent Research Findings

The collection of research studies using data provided by the NCERDC is outstanding in its breadth and quality. The studies listed below are a few examples of research that addresses the need for information on selected topics related to impacts of educational reforms. Similar to most research supported by the NCERDC, these studies would not be possible without the data files that we provided. Each of these studies requires that information on students be linked over time and across grade levels, and each study requires that student records are matched across databases (e.g., classroom reports and test scores), or across data sources (external data linked to NCERDC files).

a. Policy Implications for High-Risk Students

Miranda Marie Lynn, Dohyeong Kim , Jerome Reiter, Alicia Overstreet Galeano and Pamela Maxson. 2009. “Environmental Contributors to the Achievement Gap” *Neurotoxicology* 30(6): 1019–1024.

This study examines how environmental exposures may contribute to the observed minority achievement gap, with a focus on the relationship of childhood lead exposure to a number of adverse cognitive outcomes. In this analysis, data from the North Carolina Childhood Lead Poisoning Prevention Program surveillance registry were linked to educational outcomes available through the North Carolina Education Research Data Center for all 100 counties in NC. The findings confirm that there are differences in the impact of lead across the EOG distribution, and elucidate the impact of cumulative childhood social and environmental stress on educational outcomes. The effects of environmental and social stressors (especially as they stretch out the lower tail of the EOG distribution) demonstrate the particular vulnerabilities of socioeconomically and environmentally disadvantaged children. Given the higher average lead exposure experienced by African American children in the United States, lead does in fact explain part of the achievement gap.

Fletcher, J. 2010. Spillover effects of inclusion of classmates with emotional problems on test scores in early elementary school. *Journal of Policy Analysis and Management*, 29: 69–83.

Over the last decade, the federal government has directed schools to provide educational instruction for students with special needs in general education settings to the extent possible. While there is mixed evidence on the effects of these inclusion policies on the students with special needs, research examining potential spillovers of inclusion on non-disabled classmates has been scarce. There is particularly little research on the effects of inclusion policies on classmates during early elementary grades. This paper begins to fill in this gap by using a nationally representative, longitudinal survey of kindergartners. Cross-sectional results suggest that having a classmate with an emotional problem decreases reading and math scores at the end of kindergarten and first grade by over 10 percent of a standard deviation, which is one-third to one-half of the minority test score gap. To control for nonrandom sorting of students to schools, as well as students to classrooms, this paper uses school-level and then student-level fixed effects. Results from the preferred empirical models suggest a decrease of approximately 5 percent of a standard deviation in math and reading scores, though the reading results are less robust. The results also indicate moderate racial and gender differences in the effects

b. Impact of Accountability Policies on Student Achievement

Ladd Helen F. and Douglas L. Lauen. 2010. “Status versus Growth: The Distributional Effects of School Accountability Policies.” *Journal of Policy Analysis and Management* 29(3): 426-450.

Although the federal No Child Left Behind program judges the effectiveness of schools based on their students’ achievement status, many policy analysts argue that schools should be measured, instead, by their students’ achievement growth. Using a ten-year student-level panel dataset from North Carolina, the study examines how school-specific pressure associated with the two approaches to school accountability affects student achievement at different points in the prior-year achievement distribution. Achievement gains for students below the proficiency cut point emerge in response to both types of accountability systems. In contrast to prior research highlighting the possibility of educational triage, these studies indicate little evidence that schools in North Carolina ignore the students far below proficiency under either approach. An important finding is that the status, but not the growth, approach reduces the reading

achievement of higher performing students, with the losses in the aggregate exceeding the gains at the bottom. These analyses suggest that the distributional effects of accountability pressure depend not only on the type of pressure for which schools are held accountable (status or growth), but also the tested subject.

Ahn, Thomas. 2011. “Optimal Matching of Teachers and Schools under Accountability Pressure.” (University of Kentucky, working paper)

Accountability systems are designed to introduce market pressures to increase efficiency in education. One potential channel by which schools may increase efficiency is to recruit effective teachers in the transfer market. I use a maximum score estimator model, North Carolina public school data, and the state’s unique accountability system to analyze how schools respond to accountability pressure in the teacher transfer market. Results show that schools under a high degree of accountability pressure will match with teachers who are proven to be effective in raising test scores, while ignoring teachers with observable measures of ‘expertise,’ such as certification. Accountability pressure seems to motivate schools to compete against high achieving schools for effective teachers (and succeed).

c. Schooling and Life-Course Outcomes

Clotfelter, Charles, Helen Ladd, Clara Muschkin, and Jacob Vigdor. (Duke University, ongoing) “Schooling and Life-Course Outcomes in Early Adulthood: Enrollment and Success in Community Colleges.” Project funded by the Smith Richardson Foundation and the Sanford School of Public Policy.

This study originated with Duke University’s Beyond Test Scores Working Group, which explores important youth transitions beyond the public schools using data that link students’ school experience with later choices and outcomes. This study uses data from the North Carolina Community College System (NCCCS), linking records for community college students to their public school records. The research is currently addressing two main questions. First, who attends community colleges? Second, what factors are associated with the rate of progress through a community college and obtaining an associate’s degree?

Ladd, Helen, Kenneth Dodge, and Clara Muschkin. (Duke University, ongoing) From Birth to School: Early Childhood Programs and Educational Outcomes in North Carolina. Project funded by the Smith Richardson Foundation, CALDER, and the Sanford School of Public Policy.

This study is part of a larger research effort focusing on the influence of early childhood factors on performance in school. The goal of this study is to understand the student-level effects of access to two types of state programs that provide funding to NC counties for early childhood initiatives, including preschool. We use rich administrative data at the level of the individual to track children longitudinally from birth to third grade. We examine the progress of multiple cohorts of students, with variation by year and county in the level of exposure to two early childhood programs. Preliminary findings indicate that county-level investments in each

program have positive and significant effects on students' third grade test scores, and on lowering special education placements and grade retentions as of the third grade.

Dissemination

The North Carolina Education Research Data Center has become a leading model for working with state agencies to share administrative data and manage information so that many people can conduct research. Data sharing and dissemination of research findings are important components of the Data Center's mission.

As is evidenced in Appendix D, researchers supported by the NCERDC have presented and published their findings extensively in academic and policy conferences such as the annual meetings of the American Education Finance Association, the Association for Public Policy and Management, and the American Educational Research Association. In addition, the NCERDC provides opportunities for sharing and disseminating research. NCERDC-related studies inform policy directly, because researchers make their findings available to education administrators and policymakers through presentations, policy briefs, and efforts to disseminate information in a variety of media forms. The NCERDC has organized annual conferences with officials from the state education agency, specifically for sharing key research findings and identifying priority areas of inquiry. NCERDC will continue to support these efforts because they result in a fruitful exchange of ideas and facilitate the translation of the implications of academic research for policy and practice.

The NCERDC maintains a website as a tool for disseminating research findings as well as information about Data Center activities. This website provides agency personnel and academic researchers with continually updated information on available data and on research supported by the Data Center (www.nceddatacenter.com).

The Data Center provides regular updates to the North Carolina DPI so that policymakers know of the studies conducted. The NCERDC supports the exchange of ideas and information between researchers and policymakers by providing the state with a lay summary of each study that uses the education data files. Scholars often need help with translating findings into user-friendly formats. The NCERDC helps scholars write summaries and then disseminates them to state officials.

Dr. Dodge, Dr. Muschkin, and Ms. Bonneau are actively involved with several initiatives to share data across North Carolina agencies, for the purpose of enhancing education research and informing education policy. These initiatives include:

- Data sharing efforts among agencies involved in Early Childhood Education.
- Collaborative efforts with the NC Community College system to link school and post-secondary data.
- Active involvement with an important project for enhancing data collection processes at the NC Department of Public Instruction. The NCDPI was a 2007 recipient of a five-year US Department of Education statewide longitudinal data systems grant for enhanced management, disaggregation, and analysis of student information for

accountability reporting purposes. Dr. Dodge and Dr. Muschkin serve on this project's External Stakeholders Advisory Committee.

- Kara Bonneau is a member of the NC Action for Children Data Advisory Council, which promotes data sharing across state agencies.
- The NCERDC was actively involved in the 2010 Family Impact Seminar, a program organized by Duke University for NC legislators to learn about issues, policies, and programs addressing school suspension. Kara Bonneau assisted in preparing seminar documents, and Clara Muschkin presented an overview of trends and related findings from research using NCERDC data.