Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers

Raquel Bernal, *Universidad de los Andes Colombia*

Michael P. Keane, *University of New South Wales and Arizona State University*

We evaluate the effect of child care versus maternal time inputs on child cognitive development using single mothers from the NLSY79. To deal with nonrandom selection of children into child care, we exploit the exogenous variation in welfare policy rules facing single mothers. In particular, the 1996 welfare reform and earlier state-level policy changes generated substantial increases in their work/child care use. We construct a comprehensive set of welfare policy variables and use them as instruments to estimate child cognitive ability production functions. In our baseline specification, we estimate that a year of child care reduces child test scores by 2.1%.

I. Introduction

The effect of parental time inputs and child care use (and/or quality) on child development has been widely analyzed, especially in the psychology and sociology literatures. Economists have also recognized the importance of this issue. Some recent studies find that the factors determining individuals' labor market success are already largely in place by

We acknowledge support from Australian Research Council grants FF0561843 and DP0774247, and Raquel Bernal acknowledges financial support from the Searle Fund. Contact the corresponding author, Raquel Bernal, at rbernal@uniandes.edu.co.

[*Journal of Labor Economics*, 2011, vol. 29, no. 3]

0734-306X/2011/2903-0002$10.00

about age 16.[1] Thus, policies to enhance human capital at later ages (e.g., college tuition subsidies) have, at best, a minor impact. Naturally, such findings focus attention on human capital development at early ages, including the role of child care. In this article, we take a small step toward learning more about development of cognitive ability at very young ages (i.e., up until age 6).

Prior research has shown that children's achievement as early as age 7 is a strong predictor of a variety of later life outcomes. We provide new evidence of a strong association between test scores at ages 4–6 and completed education, even conditional on a rich set of family background controls. Thus, the issue of what determines cognitive ability at early ages appears to be critical for the design of public policy aimed at improving labor market outcomes. Unfortunately, results from the previous literature on determinants of children's cognitive achievement are inconclusive at best.

A major challenge in estimating determinants of achievement is that available data are often deficient. For example, inherited ability cannot be perfectly measured, creating difficult problems of endogeneity and self-selection. In fact, a key reason for the diverse results of the previous literature may be the failure of many studies to adequately control for biases arising from two factors: (i) women who work/use child care may be systematically different from those who do not, both in the constraints they face and their tastes; (ii) the child's cognitive ability, which is at least partially unobserved by the econometrician, may itself influence the mother's decisions. In general, mothers' work and child care decisions may depend on unobserved characteristics of both mothers and children.

To clarify the problem, consider two example cases. In case 1, women with higher skills are more likely to have children with high cognitive ability endowments and are also more likely to work/use child care. Failure to control for this correlation would cause estimated effects of maternal employment (child care) on child cognitive outcomes to be biased upward. In case 2, mothers of low-ability-endowment children may compensate by spending more time with them. Then, mothers of high-ability children are more likely to work (use child care). Again, the estimated effect of maternal employment (child care) on cognitive outcomes is upwardly biased. Clearly, such sample selection issues make evaluating the effects of women's decisions on child outcomes very difficult.

In this article, we estimate child cognitive ability production functions for single mothers in the National Longitudinal Survey of Youth 79 (NLSY79). We focus on single mothers because major changes in welfare rules in the 1990s led to dramatic and plausibly exogenous variation in

[1] See, e.g., Keane and Wolpin (1997, 2001), Cameron and Heckman (1998), and Cunha and Heckman (2008).

the work incentives of single mothers. From 1993 to 1996, 43 states received federal waivers authorizing state-level welfare reform. Then, in 1996, the federal Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) implemented substantial rule changes nationally, giving states much more leeway to set local rules. Major rule changes involved benefit and work requirement time limits, earnings disregards, and child care assistance. These policy changes greatly increased employment and child care use among single mothers with children aged 0–5. Indeed, the percentage of single mothers working increased from 59% in 1992 to 78% in 2001.

Thus, we construct an extensive set of state/time-specific welfare rules, as well as local demand conditions, and we use these as instruments in the estimation of the cognitive ability production function. One key source of identification comes from comparing outcomes for children born before 1990 versus those born later since waivers and PRWORA only affected the latter group. However, benefit levels and local demand conditions—which also have good explanatory power for the behavior of single mothers—differ substantially over states and over time for the whole sample period.

An important technical problem arises because the welfare rules are very complex. Thus, we need many variables to characterize them. As a result, we face a "many-instruments problem." That is, two-stage least squares (2SLS) estimates can be severely biased (toward ordinary least squares [OLS]) when the number of overidentifying instruments is large. (See, e.g., Stock and Yogo 2004; Anderson, Kunitomo, and Matsushita 2005; Andrews and Stock 2006; and Hansen, Hausman, and Newey 2008.) We deal with this problem in two ways. First, we use the limited information maximum likelihood (LIML) estimator, which corrects the 2SLS bias in the many-instruments case (see Hansen et al. 2008). Our 2SLS estimates fall in between those of LIML and OLS, suggesting that 2SLS does suffer from severe bias. Stock and Yogo provide a test for whether the many-instruments problem induces biases (in estimates or test sizes) that are large relative to the OLS bias. The Stock-Yogo test suggests a many-instruments problem for 2SLS, but it gives no evidence of a problem for LIML.

Second, we use factor analysis to reduce the size of the instrument set. Using only the most important factors as instruments, we obtain 2SLS results very similar to LIML results using the full instrument set. Using factors as instruments also increases efficiency relative to using LIML alone.

The estimates of our baseline specification imply that 1 additional year of child care use reduces cognitive achievement test scores by 2.1%. This corresponds to 0.114 standard deviations, so it is a substantial effect. This result is quite robust in that it differs only modestly across a wide range of production function specifications, instrument sets, and samples. The robustness to the instrument set is comforting since it is well known that in general the instrumental variable (IV) estimator only estimates a local average treatment effect, so in principle IV estimates may vary greatly depending on the instruments used.

However, this general finding of a negative effect masks important differences across child care types, maternal education, and child gender. In particular, we find that formal center-based care has no adverse effect on cognitive achievement. Only informal care (i.e., non-center-based care by grandparents, siblings, other relatives, or nonrelatives) has significant adverse effects. We estimate that an additional year of informal care causes a 2.6% reduction in test scores. Our overall negative estimate of the effect of child care obtains because 75% of single mothers use informal care.

Finally, it is interesting to examine how the welfare policy changes of the mid-1990s affected the test scores of children of single mothers. Reduced form estimates (i.e., substitute the welfare rules for the endogenous variables in the outcome equation) imply that test scores were modestly reduced.

The article is organized as follows. In Section II, we review the relevant literature. In Section III, we describe the welfare policy and local demand variables that we use as instruments. Section IV derives the cognitive ability production function that we estimate. In Section V, we describe the data, and in Section VI, we present the estimation results. Section VII concludes.

II. Literature Review

A. The Effect of Maternal Employment and Child Care on Children's Cognitive Outcomes

Many prior studies, mostly in the developmental psychology literature, used the NLSY to assess effects of maternal employment and child care use on child cognitive development. Recent reviews of this literature include Haveman and Wolfe (1995), Lamb (1996), Love, Schochet, and Meckstroth (1996), Blau (1999a), Ruhm (2002), and Blau and Currie (2004). Here we briefly review the most relevant aspects of the literature. A more detailed review is provided in the online version of this article in appendix D. The most obvious feature of the existing literature is that it has produced very mixed results. Regarding the effects of maternal employment on child outcomes, about a third of the studies report positive effects, a third report negative effects, and the rest report ambiguous effects. Results on effects of child care are similarly mixed.

Reasons for this diversity of results include data limitations, as well as the wide range of specifications and estimation methods used. To see the problems that researchers in this area face, consider the following equation, interpretable as a cognitive ability production function:

$$\ln S_{ijt} = \alpha_1 T_{ijt} + \alpha_2 C_{ijt} + \alpha_3 G_{ijt} + \alpha_4 Z_{ijt} + \mu_j + \delta_{ij} + \varepsilon_{ijt}. \qquad (1)$$

Here S_{ijt} is a cognitive outcome (i.e., test score) for child i of mother j at age t. The log is typically taken since test scores are positive. Variable T_{ijt} is a measure of the maternal time input up through age t. This might be a scalar, as in a specification where only cumulative or current inputs matter, or a vector, if inputs at different ages have different effects. Similarly, C_{ijt} is a measure of nonmaternal time inputs (i.e., child care), and G_{ijt} represents goods and service inputs. Next, Z_{ijt} is a set of controls for the child's initial (or inherited) ability endowment, for example, the mother's education, the AFQT (Armed Forces Qualification Test) score, or the child's birth weight. The error components, μ_j and δ_{ij}, are family and child effects that capture parts of the child's unobserved ability endowment. Finally, ε_{ijt} is a transitory error term that captures measurement error in the test along with shocks to the child development path.

While this general setup underlies, at least implicitly, most papers in the literature, none can actually estimate equation (1). One key problem is that the maternal time input T and the goods inputs G are not directly observed. Most papers ignore the problem that T is unobserved and simply use maternal employment or child care use in place of maternal time. Similarly, most papers simply ignore G, while a few proxy for it using household income or the NLSY's "HOME" environment index. The latter is problematic since it is based not just on goods inputs (e.g., books in the home) but also on time inputs (e.g., time spent reading to the child). A second problem is that many papers estimate specifications that include only current inputs, which is obviously a very strong assumption.

The third major problem is that most papers estimate equation (1) by OLS, ignoring the potential endogeneity of inputs, that is, correlation of maternal work and child care use decisions, as well as goods inputs, with the unobserved ability endowments μ_j and δ_{ij}. A few recent studies have tried to overcome this problem by (*a*) using an extensive set of controls for skill endowments, (*b*) using child or family fixed effects, or "value-added" models,[2] and/or (*c*) using instrumental variables.

As we note in online appendix D, even studies that use extensive controls for the child's skill endowment and/or that use fixed effects produce a wide range of conflicting results. The use of child fixed effects (as in Chase-Lansdale et al. 2003) identifies maternal work or child care effects from changes in these inputs over time, but it does not account for potential endogeneity of input changes. In contrast, the use of sibling dif-

[2] In the value-added approach, the test score in period t (S_{ijt}) is a function of the outcome in period $t-1$ and the inputs in period t, the idea being that the lagged test score proxies for the child's ability at the start of a period.

ferences (as in James-Burdumy 2005) eliminates the mother (or household) fixed effect μ_j but does not eliminate the child effect δ_{ij}. Thus, the household fixed effect (FE) estimator requires that input choices be unresponsive to the child's specific ability endowment. However, it is plausible that mothers make time compensations for children depending on the child's ability type. Also, the household FE estimator assumes that input choices do not respond to prior sibling outcomes.

Blau (1999a) and Duncan and NICHD (2003) employ value-added specifications. But, as they point out, this also does not provide a panacea for dealing with unobserved child ability. The value-added model faces the problem that estimates of lagged dependent variable models are inconsistent in the presence of fixed effects like μ_j and δ_{ij}.[3] Also, it does not deal with endogeneity arising because current inputs may respond to lagged test scores. An IV approach is necessary to deal with the endogeneity problems that the FE and value-added approaches do not address.

To our knowledge, only two prior papers used instrumental variables: Blau and Grossberg (1992) and James-Burdumy (2005). Both look at the effects of maternal work on child outcomes and do not examine the effects of child care per se. More important, their instruments are extremely weak. As a result, standard errors on the maternal work variables in their 2SLS regressions are so large that no plausibly sized effect could be significant (i.e., in each case, to attain 5% significance, maternal work over 3 years would have to change a child's tests scores by roughly 50 points, or three standard deviations). Thus, their attempts to use IV were not successful.[4] The main advantage of our approach is that the instruments we employ are much stronger, as we will see in Section VI.

Bernal (2008) takes a different approach by estimating a structural model of work and child care choices of married women. She estimates a child cognitive ability production function—with mother's work and child care use as inputs—jointly with the mother's work and child care decision rules, thus implementing a selection correction. Her results suggest that 1 full year of maternal work and child care use causes a 1.8% reduction in test scores of children ages 3–7.

It is interesting to extend this work to single mothers for two reasons: first, to see if the results generalize; second, welfare rules have large effects on work/child care use by single mothers, so as instruments they provide

[3] Estimation of a first-differenced version of the value-added specification would eliminate the fixed effects, but Blau (1999a) points out that this is difficult or impossible due to limitations of existing data. This would require three outcome observations and two lagged input observations. Even if it were feasible, this approach would entail a severe efficiency loss.

[4] For this reason, James-Burdumy's preferred specification uses sibling differences to control for household fixed effects and does not use IV.

a strong basis for identification. It is difficult to find plausibly exogenous variables that affect the behavior of married women so strongly.

Aside from the above studies, several papers estimate cognitive ability production functions, but they only do so for children old enough to be in preschool or primary school (as opposed to child care). Currie and Thomas (1995) look at preschool inputs (i.e., Head Start); Liu, Mroz, and van der Klaauw (2003) study 5–15-year-olds; and Todd and Wolpin (2007) and Cunha and Heckman (2008) look at 6–13-year-olds. Thus, none of these studies address how child care affects child outcomes.[5]

B. Relationship between Test Scores and Subsequent Outcomes (Wages, Education, and So Forth)

Several studies examine the relation between childhood test scores and subsequent outcomes such as education and wages. This research finds that children's test scores are strong predictors of a variety of outcomes in later life. The studies linking test scores at the earliest ages to later outcomes use the British National Child Development Study. Connolly, Micklewright, and Nickell (1992) find a significant positive relationship between scores at age 7 and earnings at age 23. Harmon and Walker (1998) and Robertson and Symons (2003) find a positive association between scores at age 7 and earnings at age 33. Hutchinson, Prosser, and Wedge (1979) and Currie and Thomas (2001) find that test scores at age 7 are good predictors of scores at age 16.

These prior studies, however, look at test scores at age 7 or older,[6] while we study test scores at ages 3–6. Do tests at these early ages still predict subsequent achievement? In table A1 in appendix A, we present evidence from the NLSY that PPVT (Peabody Picture Vocabulary Test) scores at age 4 and PIAT (Peabody Individual Achievement Test) scores at ages 5–6 are significantly related with educational attainment. For example, consider a one-point increase in the math score at age 6 (roughly a 1% increase since the mean is 99.7). Holding background variables such as mother's AFQT and education fixed, this is associated with an increase in education

[5] The first three papers adopt conventional empirical approaches (sibling fixed effects, a structural model, and a value-added model, respectively). Cunha and Heckman (2008) adopt the novel approach of treating investment in children as a latent variable. The items of the NLSY "HOME" environment index are used as noisy measures of investment. This helps deal with the widespread problem of missing data on key inputs that plagues this literature. A limitation is that, having estimated the effect of investment, it is not clear how observables like parental time or income map into the level of investment.

[6] The studies we are aware of that use U.S. data (i.e., Murnane, Willet, and Levy 1995; Neal and Johnson 1996; and Zax and Rees 1998) look at test scores measured at age 14 or later. (See app. D [available in the online version of this article] for details.)

(measured at age 18 or later) of .019 years. Similarly, a 1% increase in the reading score at age 6 is associated with increased schooling of .025 years. These estimated impacts are fairly substantial since our estimates imply that a year of informal child care reduces test scores by 2.6%. This translates into a drop in completed schooling of .050 to .065 years, which is a large effect.[7] Also, a striking aspect of the results is that mother's AFQT is not a significant predictor of completed education. Thus, child test scores, even at ages 4–6, are better predictors of later outcomes than are mother's scores (see online app. D for further details).

III. Construction of Instruments Using Welfare Rules and Other Policy Variables

To deal with the endogeneity of maternal work/child care, we use welfare rules as instruments to estimate cognitive ability production functions for children of single mothers. These rules are known to have a large impact on the labor supply of these mothers (see Moffitt 1992). To construct our instruments, we collected information on state welfare policies from many sources (see Fang and Keane [2004] and Bernal and Keane [2010] for details). Here we discuss the key aspects of the welfare rules that are relevant to this work. Table 1 presents the complete instrument list. Each instrument has up to three subscripts: i for individual, s for state, and t for period (quarter).

A. Benefit Termination Time Limits

Under Aid to Families with Dependent Children (AFDC), single mothers with children under age 18 could receive benefits as long as they met income and asset screens. However, under Section 1115 Waivers and Temporary Aid for Needy Families (TANF), states could set time limits on benefits. There is substantial heterogeneity across the states in how the limits were set. For instance, California sets a 5-year limit, while Texas and Florida set limits in the 2–3 year range.

The direct effect of time limits is simply that one becomes ineligible on hitting the limit. The indirect effect is that women may "bank" months of eligibility for later use. We use eight variables to capture these effects (see table 1). These include, for example, a dummy for whether a state had imposed a time limit (TLI_{st}) by time t, a dummy indicating if a time limit could be binding for a particular woman ($\mathrm{TL\_HIT}_{ist}$), and

[7] The following back-of-the-envelope calculation helps put these figures in perspective. Say that people are of two types, those who finish high school (12 years) and those who finish college (16 years), and that 20% finish college. To increase average completed schooling by .06 years, the percentage finishing college must increase to 21.5%, a 7.5% increase.

a dummy for a particular woman's maximum potential remaining eligibility ($\text{REMAIN\_TL\_ELIG}_{ist}$).

Some of our instruments are at the state level, and some are person specific. For example, consider TL_HIT_{ist}. Say a woman resides in a state that imposed a 5-year time limit 6 years earlier. Then it is possible that she could have hit the limit, but this is so only if her oldest child is at least age 5. If not, she could not yet have participated in TANF for 5 years. Crucially, however, we never use a woman's actual welfare participation history to determine if she had hit a time limit since the actual history is endogenous. Our person-specific instruments are functions only of policy rules and child ages, which we assume are exogenous (conditional on the child and mother age controls in the main equation).

B. Work Requirement Time Limits and Work Requirement Exemptions

The TANF recipients must commence "work activities" within 2 years to continue receiving benefits. But many states adopted shorter time limits, and some states exempt parents with young children. Thus, within a state, there is variation across women in whether work requirements can be binding, based on when the state implemented TANF, the length of the time limit, and child ages. We constructed a total of nine variables, listed in table 1, to capture these various effects. For example, WR_HIT_{ist} indicates whether a woman could potentially have been subject to work requirements, and CHILD_EXEM_{st} is a dummy for whether state s had a young child exemption in place at time t.

C. AFDC/TANF Benefit Levels, Earnings Disregards, and Benefit Reduction Rates

The AFDC/TANF benefits are determined by a state-specific grant level, increasing in number of children, that is reduced by a fraction of recipient earnings. We use four variables to characterize the system: the (real) state/time-specific grant levels for families with one and two children ($\text{BEN}(1)_{st}$ and $\text{BEN}(2)_{st}$); the "benefit reduction rate," which is the rate benefits are reduced with earnings ($\text{PERC\_DISREGARD}_{st}$); and the "earnings disregard," a part of earnings that is disregarded before calculating any benefit reduction ($\text{FLAT\_DISREGARD}_{st}$). Note that in constructing grant levels we do not condition on actual family size since we treat fertility as endogenous. Grant levels have always differed greatly by state, but TANF created substantial state heterogeneity in disregards as well.

Table 1
List of Instruments

Variable	Description
Time limits (TL):	
TLI_{st}	Dummy for whether state s has time limit in place in period t
TL_LENGTH_{st}	Length of time limit in state s in period t
$ELAPSED\_TL_{st}$	Time (in months) elapsed since the implementation of time limit in state s
TL_HIT_{ist}	Dummy variable indicating whether a woman could have hit time limit
$ELAPSED\_TL\_HIT_{ist}$	Time elapsed since woman i may potentially be subject to time limit
$REMAIN\_TL\_ELIG_{ist}$	Maximum potential remaining length of a woman's time limit, constructed as: $TL\_LENGTH_{st} - \min\{AGE\_OLDEST\_CHILD_{ist}, ELAPSED\_TL_{st}\}$
$REMAIN\_ELIG_{ist}$	Remaining length of time to be categorically eligible for welfare benefits: $18 - AGE\_YOUNGEST\_CHILD_{ist}$
$DCHILDBEN_{st}$	Dummy variable indicating whether the child portion of the welfare benefit continues after time limit exhaustion
Work requirements (WR):	
DWR_{st}	Dummy for whether state s has work requirement in place in period t
WR_LENGTH_{st}	Length (in months) of work requirement limit in state s in period t
$ELAPSED\_WR_{st}$	Time (in months) elapsed since the implementation of work requirement in state s
WR_HIT_{ist}	Indicator for whether a woman could be subject to a work requirement: 1 if $[WR\_LENGTH_{st} \leq \min\{AGE\_OLDEST\_CHILD_{ist}, ELAPSED\_WR_{st}\}$ and $AGE\_YOUNGEST\_CHILD_{ist} \geq AGE\_CHILD\_EXEM_{st}]$
$ELAPSED\_WR\_HIT_{ist}$	Time elapsed since woman i may be potentially subject to work requirement
$CHILD\_EXEM_{st}$	Dummy for whether state s has age of youngest child exemption in place at t
AGE_EXEM_{st}	Age of youngest child below which the mother will be exempted from work requirement in state s at time t
$WR\_ULT\_SANC_{st}$	Dummy for whether state s has a full sanction for noncompliance with work requirement at time t
$EXEMP_{st}$	Number of work requirement exemptions in state s
FDC/TANF benefits (BEN):*	
$BEN(1)_{st}$	Real AFDC/TANF maximum benefits for a family with one child
$BEN(2)_{st}$	Real AFDC/TANF maximum benefits for a family with two children
Earnings disregards (ED):	
$FLAT\_DISREGARD_{st}$	Flat amount of earnings disregarded in calculating the benefit amount.
$PERC\_DISREGARD_{st}$	Benefit reduction rate (does not include phase out)
Other policy variables (OP):	
$EITC(1)_{st}$	EITC phase in rate constructed from both the federal and state level for a family with one child
$EITC(2)_{st}$	EITC phase in rate for a family with two children
$CHILDCARE_{st}$	CCDF expenditure per single mother in state s at time t

Table 1 (*Continued*)

Variable	Description
$ENFORCE_{st}$	Child support enforcement expenditure in state *s* at year *t* per single mother
Local demand conditions (LDC):	
UE_{st}	Unemployment rate in state *s* in period *t*
$SWAGE_{st}$	Hourly wage rate at the 20th percentile of the wage distribution in state *s* in period *t*

NOTE.—The instruments used in our baseline specification also include the policy variables listed in this table and these variables interacted with mother's education and AFQT score. In addition, workbef, EXPBEF, urban, and age of mother (see definitions in table 2) are interacted with child's age.

* Benefits are put in real terms using metropolitan area and regional consumer price indices (CPIs) computed by the Bureau of Labor Statistics, given the Standard Metropolitan Statistical Area (SMSA) and state of residence of each woman in the sample.

D. Child Support Enforcement (CSE) and the Child Care Development Fund (CCDF)

Child support is an important source of income for single mothers (6.5% in the March 2002 CPS). The CSE program helps to locate absent parents and to establish paternity. The CSE expenditures increased from \$2.9 billion in 1996 to \$5.1 billion in 2002, increasing the likelihood that a single woman could collect child support. We measure state-level CSE activity by dividing state CSE expenditures by the state population of single mothers ($ENFORCE_{st}$). The CCDF helps states provide subsidized child care for low-income families. But the states can design their own programs, so a great deal of heterogeneity has emerged. We use the state CCDF expenditure per single mother ($CHILDCARE_{st}$) to measure the availability and generosity of child care subsidies in a state.

E. Other Instruments: Earned Income Tax Credit (EITC) and Local Demand Conditions

The EITC is a program that supplements wages for low-income working families. It was originally a minor program, but the wage subsidy was greatly expanded in the period 1994–96. The EITC subsidy rate varies by family size.[8] We use as instruments the wage subsidy rates for families with one and two-plus children ($EITC(1)_{st}$ and $EITC(2)_{st}$, respectively), using federal and state EITC rules.[9] As with benefit levels (see Sec. III.C), we do not condition on actual family size, which we view as endogenous. Finally, we also use two measures of local demand conditions as instruments: the state unemployment rate and the 20th percentile wage rate in the woman's state of residence at time *t*.

[8] For example, in 2003, the subsidy rates for families with one and two children were 34% and 40%, respectively.

[9] As of 2003, 17 states had enacted state earned income tax credits that supplement the federal credit.

IV. The Child's Cognitive Ability Production Function

Following Leibowitz (1974), we use the human capital production framework (see Ben-Porath 1967) to examine investments in children. Letting A_{it} be child i's (latent) cognitive ability t periods after birth, we can write the child cognitive ability production function:

$$\ln A_{it} = A(\tilde{T}_{it}, \tilde{G}_{it}, \tilde{C}_{it}, \omega_i), \tag{2}$$

where $\tilde{T}_{it}$, $\tilde{G}_{it}$, and$\tilde{C}_{it}$ are vectors of period-by-period inputs of maternal time, goods, and child care time, respectively, up through period t, and ω_i is the child's ability endowment. Goods inputs may include nutrition, books, and toys that enhance cognitive development, and so forth. Child care inputs capture contributions of alternative care providers' time to child cognitive development. These may be more or less effective than the mother's own time. Also, care in a group setting may contribute to child development by stimulating interaction with other children, learning activities at preschool, and so forth.

Several difficult issues arise in estimation of (2). A completely general specification, where the inputs and the ability endowment ω_i have different effects at each age, is infeasible due to proliferation of parameters.[10] Thus, we must restrict how inputs enter (2). Two simplifications, familiar from the human capital literature, are to assume that (i) only cumulative inputs matter and that (ii) the ability endowment has a constant effect over time. Letting $\breve{X}_{it} = \sum_{\tau=1,t} X_{i\tau}$ be the cumulative input of X up through t, and assuming inputs affect $\ln A_{it}$ linearly, we obtain:[11]

$$\ln A_{it} = \alpha_0 + \alpha_1 \breve{T}_{it} + \alpha_2 \breve{C}_{it} + \alpha_3 \ln \breve{G}_{it} + \omega_i. \tag{3}$$

Next, we further assume that the ability endowment ω_i is given by the equation

$$\omega_i = \beta_0 + \beta_1 Z_i + \hat{\omega}_i, \tag{4}$$

where Z_i is a vector of mother/child characteristics (e.g., mother's AFQT) that may be correlated with the child's ability endowment and $\hat{\omega}_i$ is the part of the endowment that is mean independent of observed characteristics. A detailed description of the variables in Z_i can be found in table 2.

Another key problem is the measurement of maternal time. This can take various forms (e.g., "quality" time vs. the child watching TV while the mother does housework), but we do not observe these distinctions in the data. Thus, we distinguish only two types of time, time with the

[10] For instance, if the effect of just one input is allowed to differ between every pair of input and output periods t and t' and we examine outcomes for 20 quarters after birth, we obtain $(20 \times 21)/2 = 210$ parameters for that input alone.

[11] It is convenient to let the cumulative goods input enter in log form for reasons that will become apparent below.

Table 2
Control Variables in the Cognitive Ability Production Function

Variable	Description
Baseline specification:	
$I[AGE_i < 20]$	Dummy for whether mother is younger than 20 years old
$I[AGE_i \geq 33]$	Dummy for whether mother is older than 32 years old
$EDUC_i$	Mother's educational attainment at childbirth
$AFQT_i$	Mother's AFQT score
$AFQT_i \times dPPVT$	Mother's AFQT score interacted with a dummy for PPVT test
$I[WORK\_BEF]_i$	Dummy for whether mother worked prior to childbirth
$I[WORK\_BEF]_i \times SKILL_i$	Work dummy interacted with mother's skill*
$EXPBEF_i$	Mother's total work experience (in number of years) prior to childbirth
$EXPBEF_i \times age_i$	EXPBEF interacted with mother's age
$MARSTAT_i$	Indicators for whether never-married, separated, widowed, or divorced
$URBAN_i$	Urban/rural residence at time of child's test
$NUMCHILD_i$	Number of children
$RACE_i$	Child's race (1 if black/Hispanic, 0 otherwise)
$RACE_i \times dPPVT$	Child's race interacted with a dummy for PPVT test
$RACE_i \times dMATH$	Child's race interacted with a dummy for PIAT-Math test
$GENDER_i$	Child's gender (1 if male, 0 if female)
BW_i	Child's birthweight
$AGECHILD_i$	Child's age at assessment date
$dPPVT_i$	Dummy for whether the corresponding test is PPVT
$dMATH_i$	Dummy for whether the corresponding test is PIAT-MATH
Alternative specifications also include:	
AGE_i	Age of the mother at childbirth
$(AGE_i)^2$	Age of the mother at childbirth squared
$NUMCHILD_{0-5\ i}$	Number of children 0–5 years of age
$NUMCHILD_{6-17\ i}$	Number of children 6–17 years of age
$C_{it} \times EDUC_i$	Cumulative child care use interacted with mother's education
$C_{it} \times AFQT_i$	Cumulative child care use interacted with mother's AFQT score
$C_{it} \times NUMCHILD_i$	Cumulative child care use interacted with number of children
$C_{it} \times NONWHITE_i$	Cumulative child care use interacted with nonwhite dummy
$C_{it} \times BLACK_i$	Cumulative child care use interacted with black dummy
$C_{it} \times HISP_i$	Cumulative child care use interacted with Hispanic dummy
TIME	Calendar time trend (0–16 starting in year 1984)

* The variable "skill" is defined as the residual from a regression of mother's initial wage on age, age squared, education, and race.

mother versus time in child care, and we assume that $T_{it} = T - C_{it}$, where T is total time in a period. Then, we can rewrite (3) as

$$\ln A_{it} = \alpha_0 + (\alpha_1 T)t + (\alpha_2 - \alpha_1)\breve{C}_{it} + \alpha_3 \ln \breve{G}_{it} + \omega_i. \tag{5}$$

Thus, we can only estimate $\alpha_2 - \alpha_1$, the effect of time in child care relative to that of mother's time.

The next problem is that goods inputs G_{it} are largely unobserved. For example, the NLSY contains information on books in the home but little about nutrition, health care, tutors, and such. Thus, we proxy for the (log) cumulative goods input using the demand for goods, conditional on cumulative income since childbirth, mother/child characteristics Z_i (which affect permanent income and also preferences), the child's unobserved ability $\hat{\omega}_i$,[12] and child age. Specifically, we write

$$\ln \breve{G}_{it} = \gamma_0 + \gamma_1 Z_i + \gamma_2 \hat{\omega}_i + \gamma_3 \ln \breve{I}_{it} + \gamma_4 t + \varepsilon_i^g, \tag{6}$$

where the stochastic term ε_i^g captures the household's tastes for investment in the form of goods.[13] Note that prices of goods and child care are not included in (6) since we cannot measure them, so in effect we assume they are fixed. One reason the prior literature failed to find good instruments for child care is lack of sufficient price variation in the data.[14] Substituting (6) and (4) into (5), we obtain

$$\begin{aligned} \ln A_{it} &= \alpha_o + (\alpha_1 T)t + (\alpha_2 - \alpha_1)\breve{C}_{it} \\ &\quad + \alpha_3[\gamma_o + \gamma_1 Z_i + \gamma_2 \hat{\omega}_i + \gamma_3 \ln \breve{I}_{it} + \gamma_4 t + \varepsilon_i^g] + \beta Z_i + \hat{\omega}_i \\ &= (\alpha_o + \alpha_3 \gamma_o) + (\alpha_1 T + \alpha_3 \gamma_4)t + (\alpha_2 - \alpha_1)\breve{C}_{it} + \alpha_3 \gamma_3 \ln \breve{I}_{it} \\ &\quad + (\beta + \alpha_3 \gamma_1) Z_i + (1 + \alpha_3 \gamma_2)\hat{\omega}_i + \alpha_3 \varepsilon_i^g \\ &= \phi_o + \phi_1 t + \phi_2 \breve{C}_{it} + \phi_3 \ln \breve{I}_{it} + \phi_4 Z_i + \hat{\hat{\omega}}_i + \hat{\varepsilon}_i^g. \end{aligned} \tag{7}$$

Equation (7) is estimable since all independent variables are observed. However, endogeneity problems arise if the inputs $\breve{C}_{it}$ and $\ln \breve{I}_{it}$ are correlated with the error term, which includes both the child ability endowment $\hat{\omega}_i$ and tastes for goods investment $\hat{\varepsilon}_i^g$. This appears likely. Decisions to work and use child care may well be correlated with the child's ability endowment, and the work decision affects income as well. Furthermore, tastes for child care may well be correlated with tastes for goods invest-

[12] Child ability may matter because (i) mothers choose goods inputs based on it (e.g., buying educational toys to help a child with learning problems) or (ii) it affects the goods a child wants (e.g., a high-ability child may want more books).

[13] This would arise due to heterogeneous preferences for child quality. Variable ε_i^g may also be influenced by the child's tastes.

[14] Equation (6) is consistent with several alternative models of investment. For example, if $\gamma_3 = 0$, there is a fixed level of investment determined by permanent characteristics and the cumulative goods input grows at a rate given by γ_4. At the other extreme, if $\gamma_1 = \gamma_2 = \gamma_4 = 0$ and $\gamma_3 = 1$, then demand for goods is simply proportional to current income ($G_{it} = \exp(\gamma_0) \times I_{it}$).

ment in children.[15] Thus, estimation of (7) by OLS is unlikely to be appropriate.

In order for welfare rules and local demand conditions to be valid instruments for cumulative child care and income in (7), they must be uncorrelated with $\hat{\omega}_i$ and $\hat{\varepsilon}_i^g$. This seems like a plausible exogeneity assumption.[16] These instruments are also likely to be powerful since it is well known that welfare rules have important effects on work (and, hence, child care) decisions of single mothers.

It is important to note one limitation of our approach. Rosenzweig and Schultz (1983) refer to an equation like (7), where proxy variables are substituted for one or more unobserved inputs, as a "hybrid" production function. As long as demand for goods conditional on income and mother/child characteristics is well described by (6), estimation of (7) using appropriate instruments will identify $\alpha_2 - \alpha_1$, the effect of child care time relative to maternal time. However, suppose that demand for goods also depends on child care time, as would be the case if mothers compensate for a low time input by increasing the goods input. Then our estimate of the effect of child care would be contaminated by the effect of any change in goods inputs that she may choose as a result of using child care (holding income fixed). Appendix E in the online version of this article discusses this and other related issues in more detail. We stress, however, that the alternative of ignoring missing inputs leads to omitted variable bias. As noted by Todd and Wolpin (2007), it is not obvious a priori which approach would lead to greater bias.

Finally, we do not observe the actual cognitive ability of children, but we have available a set of (age-adjusted) cognitive ability test scores that measure ability with error (the PPVT, PIAT-math, PIAT-reading). Let S_{it} be the test score observed in period t, and assume the following measurement process:

$$\ln S_{it} = \ln A_{it} + \eta_1 d_{i1t} + \eta_2 d_{i2t} + \eta_3 d_{it} X_i + \varepsilon_{it}, \tag{8}$$

where d_{1t} and d_{2t} are cognitive ability test dummies that capture the mean differences in scores across the three tests,[17] and where ε_{it} combines measurement error and shocks to the development path. A key question is

[15] For instance, a mother with a high taste for child quality may both spend more time with the child (i.e., use less child care) and invest more in the child in the form of goods. This would tend to bias estimated effects of child care usage in a negative direction since not only the maternal time input but also the goods input is lower for children in child care.

[16] One might be concerned that states that adopted stricter welfare reform tended to have relatively high/low test scores initially. But in online app. J, we present evidence that this was not the case.

[17] In particular, $d_{1t} = 1$ if S_t corresponds to the Peabody Picture Vocabulary Test (PPVT) and $d_{2t} = 1$ if S_t corresponds to the Peabody Individual Achievement Test-Math Section (PIAT-math). The PIAT-reading test is the base case.

whether it is appropriate to pool the three tests. Below we provide evidence that it is appropriate provided that we include the terms $d_{it}X_i$, where X_i is a (small) subset of the regressors whose relation to the conditional mean differs by test. By substituting (7) into (8), we obtain

$$\begin{aligned} \ln S_{it} = {} & \phi_o + \phi_1 t + \phi_2 \breve{C}_{it} + \phi_3 \ln \breve{I}_{it} + \phi_4 Z_i \\ & + \eta_1 d_{i1t} + \eta_2 d_{i2t} + \eta_3 d_{it} X_i + v_{it}, \end{aligned} \tag{9}$$

where $v_{it} = \hat{\hat{\omega}}_i + \hat{\varepsilon}_i^g + \varepsilon_{it}$. Equation (9) is the baseline specification that we estimate.

In our empirical work we consider many alternative versions of (9). We compare cumulative and current input specifications. We estimate models that allow for heterogeneous treatment effects in the form of interactions between child care use and observed characteristics of the mother (e.g., education and AFQT). We also test for differences in the effect of child care depending on characteristics of the provider (i.e., formal vs. informal) and of the child (i.e., age, race, and gender).

V. Data

A. Construction of the Sample

We use data from the National Longitudinal Survey of Youth 1979 cohort (NLSY79). Panel members were ages 14–21 on January 1, 1979, and they were interviewed annually since 1979. There is a core random sample and oversamples of blacks, Hispanics, poor whites, and the military. A survey of children of the NLSY79 female respondents was begun in 1986 (CNLSY79). It contains the cognitive ability tests that we use in our analysis. We use only single mothers in the NLSY79 precisely because their work/child care decisions were greatly affected by policy changes in our sample period.

We require that women in our sample be single (or not cohabiting with a male) during 5 years following the birth of a child and that we observe at least one test score for the child. There are 1,464 mother/child pairs in the NLSY79 who satisfy these criteria, and they had a total of 3,787 test score observations (2.59 per child). An issue we did not discuss in deriving (9) is that mothers may have multiple children, which may affect resources allocated to any one child.[18] Thus, we include the number of children in Z_i in (9). Of course, number of children may be endogenous in (9), for

[18] Of the 1,464 children in our sample, 576 are "single" children in the sense they have no siblings in the sample. (Of course, they may have older or young siblings who are too old or too young to have test scores recorded in the sample. We include such siblings when constructing the number of children in the household.) There are an additional 888 children (of 368 mothers) who do have siblings in the sample.

Table 3
Mean Characteristics of Mothers in the Sample

Description	All Mothers in NLSY	Single Mothers at Childbirth Only	Single Mothers for 5 Years after Childbirth	Our Sample
Mother's age at childbirth	24.8	23.56	23.80	23.13
	(5.56)	(5.07)	(5.15)	(4.59)
Mother's education at childbirth (in years)	12.0	11.3	11.3	11.2
	(2.475)	(1.920)	(1.917)	(1.909)
Mother's AFQT score	37.9	21.7	19.9	19.3
	(27.23)	(20.09)	(19.11)	(18.30)
Hispanic or black	.47	.73	.79	.83
	(.499)	(.445)	(.404)	(.379)
Hourly wage before childbirth (first child)*	6.32	4.74	4.90	4.39
	(7.71)	(8.23)	(9.85)	(2.01)
Total no. of children of mother	2.9	3.1	3.1	3.1
	(1.37)	(1.57)	(1.61)	(1.53)
Father present at birth	.55			
	(.497)			
N	4,814	2,528	1,820	1,464
Cases with wages at childbirth observed	2,622	1,208	753	670

NOTE.—Our sample screens are (i) that the mother does not have a husband/partner for 5 years after childbirth and (ii) that the child has at least one test score observation. Standard deviations are in parentheses. N = number of observations.
* These are 1983 dollars; these correspond to \$13.15, \$9.86, \$10.20, and \$9.14, respectively, in 2007 dollars.

example, if there is a quality/quantity trade-off, so we instrument for it using the variables in table 1.

In our sample, 251 women had children from 1990 to 2000, so waivers/TANF affected their work decisions before the children reached school age. Part of our leverage for identification comes from comparing outcomes of these children to those of 1,213 children born too early to be affected. However, even in the prereform period, some of our instruments, such as AFDC grant levels and local demand conditions, varied greatly across the states and over time, also providing an important source of identification. Also, in the postreform period, some states adopted "strict" versus "lenient" reforms.

Table 3 compares the single mothers in our sample with other mothers in the NLSY79. Note that the mothers in our sample are quite similar to the set of all single mothers in the NLSY79. So using only women who remain single for 5 years after childbirth does not appear to create a very select sample. Of course, the mothers in our sample differ substantially from typical mothers. They are younger by 1.7 years, less educated by 0.8 years, and have 30% lower wages. They are more likely to be nonwhite and less likely to work during the first year after childbirth (39% vs. 47%).

B. Measuring Maternal Time and Other Inputs and Measuring Child Cognitive Ability

Using the NLSY79 work history file, we construct an employment history for each mother in the sample for the period around the birth of each child for four quarters before birth to 20 quarters after birth (a period of 5 years). We also use retrospective data gathered in 1986, 1988, 1992, and 1994–2000 to construct quarterly child care usage histories for the first 3 years of a child's life.

Unfortunately the NLSY does not report hours of child care. It contains only an indicator for whether the mother used child care for at least 10 hours per week. This is inadequate to determine if care was full-time or part-time. However, by combining the child care variable with data on work, we can make a reasonable imputation. It is particularly helpful that for single mothers we know that child care hours must be at least as great as their work hours.[19] Thus, we use the following procedure. If a woman reports using 10+ hours per week of child care, we assume that she used child care during the quarter. If she worked full-time (i.e., 375+ hours in a quarter), we assume that child care must have been full-time, which seems clear. However, if the mother did not work (i.e., <75 hours in a quarter) but still reported using child care—not a common state for single mothers—it seems likely that child care was part-time. More difficult is making a reasonable assignment if the mother worked part-time (75–375 hours in a quarter). We decided to assume that child care was part-time in this case. We admit that this assignment is not so obvious. However, we experimented with assigning full-time child care instead and found it had almost no effect on the results.[20] Thus, we define the function:

$$I_t^c = \begin{cases} 1 & \text{if mother works full-time and used child care} \\ 0.5 & \text{if mother works part-time and used child care} \\ 0.5 & \text{if mother did not work and used child care} \\ 0 & \text{otherwise} \end{cases}$$

and form cumulative child care, $\breve{C}_t = \sum_{\tau=1}^{t} I_\tau^c t$, and current child care, $C_t = I_t^c$, where t is child age.

As we noted earlier, complete child care histories are only available for 3 years after childbirth. We impute child care choices in years 4 and 5 based on current work and work/child care histories. Again, imputation

[19] Note that we define "child care" as all nonmaternal care (whether formal care in a center or informal care, i.e., by relatives, siblings, etc.), so this statement is almost definitional. Rare exceptions would be if the woman leaves the child alone while working (in violation of the law) or is self-employed and can work at home while caring for the child. But the self-employed make up only 0.9% of the sample, and 60% of these still do use child care.

[20] This is not surprising since the cumulated child care variables constructed in these two ways have a correlation of .98.

is made easier by the fact that child care and work are so closely linked for single mothers. Thus, we set $I_t^c = 1$ or 0.5 for mothers who work full-time or part-time, respectively, in a given period t after the third year. For mothers who do not work in a given period t, we impute the child care choice using a probit model that we estimate using observed work and child care histories. As the probit coefficients in table B1 in appendix B show, child care use by nonworking mothers is very well predicted by lagged child care and work choices. For quarters in which we observe child care (i.e., the first 12), our imputations from this probit are correct in 88.7% of cases.[21]

Another input into the cognitive ability production function (9) is real household income. We measure it by summing income from all sources, including wages, public assistance, unemployment benefits, interest, dividends, pensions, rentals, alimony, child support, and/or transfers from family or relatives. Income is deflated using a region-specific CPI, just as we did for welfare benefits.

Finally we turn to the child cognitive ability measures in the CNLSY79. We use the PPVT at ages 3, 4, and 5 and the PIAT at ages 5 and 6. The PIAT consists of reading and math subtests, PIAT-R and PIAT-M. The PPVT and PIAT are among the most widely used assessments for preschool children.

C. Descriptive Statistics

Table 4 presents descriptive statistics for the analysis sample. The average log test score is 4.50, with a standard deviation of 0.186 (adjusting for mean differences across tests). Sixty-four percent of the mothers worked prior to giving birth, at an average hourly wage of $9.14 in 2007 dollars. Average annual household income is $22,700 (2007 dollars). During the 20 quarters after childbirth, mothers use 0.355 units of child care per quarter, for a total of 7.1 quarters, on average. Comparing the 1979–93 and the post-1993 periods, the child care usage rate increases 10 points (from 59% to 69%).

Figure 1 shows work and child care choices for 5 years after birth. In the first quarter, 74% of single mothers stay at home and do not use child care. The rest use child care, with 10% working full-time, 5% working part-time, and 12% staying home. By the end of 16 quarters, only 38% stay at home and do not use child care; 29% work full-time and 16% work part-time, while 17% stay home and use child care.

[21] This high degree of accuracy is not surprising given the great persistence in the data, i.e., conditional on not working at $t-1$ and t, the probability that a woman who used child care at $t-1$ continues to do so at t is 92.4%. Indeed, the unconditional persistence rate in child care use is 93.5%, while that in nonuse is 89.1%.

Table 4
Summary of Variables Used in the Empirical Analysis

Variable	Mean	SD
Log (test score)	4.49855	.1861*
Mother's education	11.208	1.8972
Mother's age	23.136	4.5820
Boys (children of single mothers)	.4976	.5001
Hispanic or black	.8262	.3790
Birth weight	111.97	21.976
Mother worked before giving birth	.6431	.4792
Wage rate prior to giving birth†	4.3938	2.0075
Accumulated work experience prior to giving birth (no. of years)	4.7202	6.0088
Never married	.7215	.4483
Separated	.1540	.3611
Divorced	.1158	.3201
Urban	.8189	.3851
Average yearly income (thousands)‡	10.9274	13.568
Cumulative income (thousands)§	51.1787	67.415
Units of child care per quarter‖	.3546	.3064
Cumulative child care use (quarters)	7.0923	6.1273
Labor force participation rate (average 1979–93)	48.23	6.5
Labor force participation rate (average 1994–99)	60.40	5.4
Welfare participation rate (average 1979–93)	58.93	4.1
Welfare participation rate (average 1994–99)	44.65	12.1
Child care use rate (average 1979–93)#	59.05	5.0
Child care use rate (average 1994–99)	69.27	5.9

* The standard deviation of log (test score) is calculated after taking out the test-specific means of the three tests, that is, the standard deviation of the residuals from a regression of log (test score) on test dummies PPVT and PIAT Math.

† 1983 dollars (values used in the regressions). This corresponds to $9.14 in 2007 dollars.

‡ Values used in the regressions are 1983 dollars. This corresponds to $22.7 thousand in 2007 dollars.

§ Values used in the regressions are 1983 dollars. This corresponds to $106.5 thousand in 2007 dollars.

‖ One quarter of full-time child care use is one unit, and one quarter of part-time child care use is one-half unit.

It is equal to one if child care is used (either full-time or part-time).

Table A2 in appendix A contains descriptive statistics for test scores. Mean scores on the PPVT, the PIAT-M, and the PIAT-R are roughly 80, 95, and 101, respectively.[22] Note that there is no clear age pattern in scores since they are age adjusted. Interestingly, score differentials between children who are white versus nonwhite and who have high school–graduate versus high school–dropout parents are already apparent in the PPVT at age 3, and there is no discernible pattern of these differentials growing over time. This highlights the importance of studying determinants of test scores at very early ages.

[22] Standard deviations seem to vary more by age than by test. For instance, at age 5, the one age where we see all three tests, the standard deviations are quite similar: 17.5, 14.3, and 15.3, respectively.

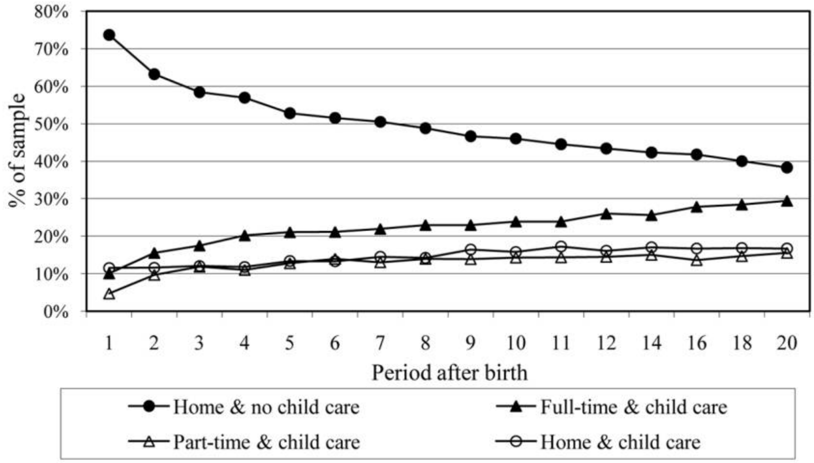

FIG. 1.—Employment and child care choices of single mothers after birth. Source: NLSY79.

VI. Estimation Results

A. The Reduced Form Regressions for the Endogenous Variables

The first stage of 2SLS, as well as the reduced form equations in LIML, use the instruments listed in table 1, along with all the exogenous variables that appear in (9)—see table 2—to predict the three endogenous variables in the model (e.g., cumulative child care, income since birth, and number of children). The procedure is complicated by the fact that the instruments are time varying and the endogenous variables are presumably functions of the instruments for all periods from birth up through time t. Thus, the set of instruments grows with t. We describe this structure in equation (C1) of appendix C.

Table 5 reports correlations of the instruments with the endogenous variables. Column 1 shows the partial correlation squared, while column 2 shows Shea's partial correlation squared.[23] For cumulative child care, these are .1735 and .1483, respectively. Column 4 shows the incremental R^2s from adding the excluded instruments. For cumulative child care, this is .0908, and the (cluster robust) F-test for joint significance of the excluded instruments is 14.74 (the 1% critical value is 1.47). These results suggest that our instruments are reasonably powerful, especially compared

[23] This partials out the correlation of an endogenous variable with fitted values of the other endogenous variables.

Table 5
Explanatory Power of the Instruments

Input	(Partial Correlation)^2 (1)	(Shea Partial Correlation)^2 (2)	R^2 with Exogenous Variables Only (3)	Incremental R^2 (4)	F-statistic* (5)	p-Value (6)
A. Endogenous variables in the baseline model:						
Cumulative child care use	.1735	.1483	.4765	.0908	14.740	.0000
Current child care use	.1314	.0986	.3378	.0807	10.970	.0000
Cumulative income	.1112	.1163	.2163	.0872	26.120	.0000
Current income	.0873	.0864	.1130	.0767	3.4200	.0000
No. of children	.3944	.3252	.2469	.2970	25.220	.0000
B. Other endogenous variables in additional models:						
Cumulative formal child care	.0943	.0997	.0906	.0858	51.470	.0000
Cumulative informal child care	.1392	.1446	.3385	.0921	16.310	.0000
Cumulative child care by nonrelatives	.0956	.1001	.0893	.0871	1.8100	.0001
Cumulative child care by relatives	.1277	.1461	.2255	.0989	14.360	.0000
Age-weighted cumulative child care	.1892	.1043	.4556	.1030	25.260	.0000

NOTE.—Instruments are variables in the main equation (see table 2); mother's age and mother's age squared; all policy variables in table 1; as well as these policy variables interacted with mother's education and mother's AFQT and child's age interacted with workbef, EXPBEF, urban, and age of mother (see definitions in table 2). Cumulative child care, cumulative income, and number of children are predicted using lags and current values of the instruments listed above. Current child care and current income are predicted using current values of the instruments listed above.

* Each F-statistic is cluster robust. Critical value at 1% is 1.47 (78 df in the numerator and 1,463 df in the denominator).

to those used in earlier attempts to study the effects of maternal employment (see Sec. II.A).

We omit the reduced form regressions to conserve on space.[24] However, it is worth noting that the 78 policy instruments have reasonable coefficients. The strongest predictors of cumulative child care are (i) if a state had a work requirement, which has a strong positive effect on work/child care use ($t = 2.6$), (ii) the number of work requirement exemptions a state allows, which has a strong negative effect ($t = -6$), (iii) the age of youngest child exemption from work requirements, which has a negative effect ($t = -2.4$), (iv) the remaining time a woman is categorically eligible for welfare, which has a negative effect ($t = -3$), (v) time elapsed since a time limit could have hit, which has a positive effect ($t = 2.5$), and (vi) time elapsed since a work requirement time limit could have hit, which has a negative effect ($t = -2.9$). Benefit levels are not individually significant (because BEN(1) and BEN(2) are highly collinear), but an F-test for their joint significant gives $p = .0000$. As expected, interactions of education with welfare policy variables are always opposite in sign to the main effects, indicating that the behavior of high-skilled mothers is less influenced by welfare rules.

Table 5 also reports Shea partial correlations for cumulative income and number of children. These are .1163 and .3252, respectively. Thus, the instruments successfully generate independent variation of the three endogenous variables. It is useful to know what instruments are important in generating this independent variation. Variables that are important for income (but not child care) are time limits, child support enforcement, and whether the woman could have potentially hit the work requirement. Variables that are important for number of children (but not child care) are whether a state has a young child exemption from the work requirement, whether the sanction for violating work requirements allows one to keep the child portion of benefits, and child support enforcement.

B. Baseline Specification of the Child Cognitive Ability Production Function

Before arriving at a baseline specification, we first test if it is appropriate to pool data from the three tests. Pooling is desirable if appropriate since it leads to an efficiency gain. Appendix F in the online version of this article contains several tests of the pooling hypothesis. Based on these tests, we cannot reject the null that the production function (9) is invariant across the three tests provided that we allow for race/test and AFQT/

[24] The reduced form first-stage regressions contain 101 variables, of which 23 are exogenous variables that also appear in the structural equation (9) and 78 are excluded instruments. As eq. (9) contains three endogenous variables, there are 75 overidentifying restrictions.

Table 6
Comparison of Results by Estimation Method

	OLS (1)	GMM (2)	2SLS (3)	FULLER[a] (4)	LIML (5)	LIML[b] (6)
Cumulative child care	.00098	−.00389*	−.00357+	−.00521+	−.00522+	−.00533*
	(.0008)	(.0015)	(.0021)	(.0028)	(.0028)	(.0025)
Log(cumulative income)	−.00324	.01142	.00742	.01035	.01037	.01062
	(.0057)	(.0087)	(.0164)	(.0242)	(.0242)	(.0266)
Mother's education	.01051*	.01130*	.01214*	.01276*	.01276*	.01297*
	(.0026)	(.0023)	(.0029)	(.0032)	(.0032)	(.0030)
Mother's AFQT	.00059*	.00057*	.00063*	.00066*	.00066*	.00066*
	(.0002)	(.0002)	(.0003)	(.0003)	(.0003)	(.0003)
R^2	.3994	.3891	.3914	.3847	.3847	.3844
k[c]				1.039	1.040	1.005
Weak/many-instruments test		5.80	5.80	5.80	5.80	15.33

NOTE.—The dependent value is log (test score). N = number of observations = 3,787. In cols. 2–5 we use as instruments the same 78 variables described in the note of table 5. Additional control variables are listed in table 2. Residuals in the score equation depart modestly from normality, exhibiting some skewness and excess kurtosis. Neither consistency nor asymptotic normality of LIML depend on normality, but small sample properties are presumably improved by the approximation being reasonably accurate. Robust standard errors (Huber-White) by child clusters are in parentheses.
[a] Fuller parameter $\alpha = 1$.
[b] Instruments are the 14 more relevant factors derived from the factor analysis of our original 78 instruments described in the note of table 5.
[c] k is the parameter of the k-class estimator, which equals one for 2SLS and exceeds one for LIML.
[+] Significant at the 10% level.
[*] Significant at the 5% level.

test interactions (i.e., let X_i in [9] include race and AFQT). Specifically, we find that nonwhites have relatively low PPVT scores relative to their PIAT scores and that AFQT has a relatively large impact on PPVT scores. There is no evidence that other parameters of (9) differ by test.

Next, we assess whether a cumulative or current child care specification is most appropriate. Appendix G in the online version of this article reports estimates of both models. Starting with the current child care specification, if we add 4 years of lagged child care, the p-value for their joint significance is .0329. Thus, we reject that only current child care matters. Also, a χ^2 test for equality of coefficients on current and lagged child care gives a p-value of .1748. This supports the cumulative specification.

Having settled on a baseline specification, our baseline results are reported in table 6. We report estimates from several methods. The OLS estimate of the effect of cumulative child care is essentially zero, and it is insignificant. However, the 2SLS estimate based on our 78 excluded instruments is −0.36% per quarter (−1.4% per year), and it is significant at the 10% level. As noted earlier, however, 2SLS based on such a large number of instruments is likely to be severely biased toward OLS. Using

LIML, we obtain an estimate of −0.52% per quarter (−2.1% per year), with a t-statistic of −1.86.

The LIML estimate implies a substantial child care effect, but it is somewhat imprecise, with a standard error one-third greater than 2SLS. Thus, we sought a way to increase precision. To do this, we factor-analyzed the 78 excluded instruments, reducing the instrument set to only 14 factors. We describe this procedure in detail in the next section. As we see in the last column of table 6, this had little effect on the LIML point estimate, but it has reduced the standard error, so the estimate is now significant at the 5% level ($t = -2.13$). We view this as our preferred estimate of the baseline model.

Our baseline model implies that a year of full-time child care reduces scores by about 2.13%. This is a substantial effect as it corresponds to $.0213/.1861 = 0.114$ standard deviations of the score distribution. Viewed another way, given our estimates in table A1, a 2.1% test score reduction at age 6 translates into a .040 to .053 year reduction in completed schooling.[25]

Table 6 also reports estimates for cumulative income, mother's education, and AFQT. The estimated effect of income since childbirth is quantitatively small and insignificant. The point estimate of .0106 in column 6 of table 6 implies that doubling cumulative income (increasing its log by .69) would increase test scores by only $(.0106)(.69) = 0.7\%$. In contrast, mother's education and AFQT are highly significant and quantitatively important. This is consistent with a view that lifetime income is more important than transitory income in determining parental investment in children and, hence, child achievement.[26] Of course, mother's education and AFQT may also be important for other reasons, including genetic transmission of ability and the fact that more educated mothers may be better at teaching children. We do not attempt to disentangle these alternative mechanisms.

C. Comparison of Alternative Estimation Methods

As noted earlier, we use the LIML estimator of Anderson and Rubin (1949) because both theory and Monte Carlo evidence suggest that it is less subject to "many-instruments" bias than 2SLS. Our results in table 6 appear consistent with this since the 2SLS estimate of the child care

[25] Repeating the calculation at the end of Sec. II, this gives about a 6% decrease in the number who attend college.

[26] This is reminiscent of findings by Cameron and Heckman (1998) and Keane and Wolpin (2001) to the effect that transitory fluctuations in parental income have little effect on college attendance decisions by youth. In addition, it is consistent with findings by Blau (1999b) and Carneiro and Heckman (2002), according to which permanent household income is significant for investments in children while transitory income is not.

effect, using 78 overidentifying instruments, is shifted about 25% of the way toward OLS.

The bottom row of table 6 reports the Cragg-Donald (1993) weak-instrument test statistic, which is 5.80 when we use all 78 instruments. Stock and Yogo (2004) develop critical values of this statistic for testing the null that the asymptotic maximal bias of 2SLS may exceed some percent of the OLS bias (under many-instrument asymptotics). With three endogenous variables and 78 excluded instruments, the critical values for the null that the 2SLS bias may exceed 20%, 10%, or 5% of the OLS bias are 5.65, 10.76, and 20.82, respectively. Thus, we cannot reject the null that the 2SLS bias may exceed 10% of the OLS bias, and we barely reject that it may exceed 20% (i.e., 5.80 vs. 5.65).

Next, consider the LIML and Fuller (1977) estimators, which give almost identical results. From Stock and Yogo (2004), the critical value to reject the null that the bias of the Fuller estimator may exceed 5% of the OLS bias is roughly 1.8 in our case—easily exceeded by our 5.80 value of the Cragg-Donald statistic. Further, the critical value for the null that bias in size for LIML test statistics may exceed 10% of the OLS bias is roughly 5.4. Thus, in contrast to 2SLS, there is no evidence of serious bias or size distortions for the LIML or Fuller estimators, even with 78 excluded instruments.

Unfortunately, the LIML and Fuller estimators result in efficiency losses (i.e., standard errors one-third greater than 2SLS). Thus, as an alternative approach to the many-instruments problem, we tried reducing the size of the instrument set. Specifically, we summarize the information contained in the 78 instruments using a smaller set of variables obtained via factor analysis. We used the principal factor method with varimax rotation. The factor scoring coefficients are calculated via the regression method. The estimated factors are linear functions of the original instruments. Thus, if the original instruments are valid, the estimated factors will be valid as well.

A common rule-of-thumb in factor analysis is to retain factors with eigenvalues greater than one, of which there are 13. However, in the present context we are not interested in obtaining a set of factors that best summarize the correlations of the 78 instruments per se. Rather, we are interested in finding the factors that best explain the endogenous variables. To do this, we regressed each of the endogenous variables on the full set of factors and retained those that were most highly significant.

For cumulative child care, the most important factors ($t > 3$), ordered by eigenvalue, are 2, 6, 8, 9, 12, 19, 21, 24, and 26. Given which variables load on each factor, we interpret them as follows. Factor 2 captures benefit levels, remaining eligibility, disregards, CSE, and EITC. Factor 6 primarily captures benefit levels. Factor 8 captures remaining eligibility, local wages, and EITC. Factor 9 captures time limits. Factor 12 captures local un-

Table 7
Explanatory Power of Instruments in First-Stage Regression for Child Care (Instruments in Table 8)

Instruments Listed in Note to Table 8	(Partial Correlation)^2	(Shea Partial Correlation)^2	Incremental R^2	F-Statistic	p-Value
All 78 instruments	.1735	.1483	.0908	14.740	.000
14 rotated factors	.1043	.0967	.0546	14.570	.000
21 principal factors	.1039	.0960	.0544	12.630	.000

NOTE.—The dependent variable is cumulative child care. R^2 of the first-stage regression with only exogenous variables = .4765.

employment. Factor 19 captures work requirements, sanctions, and strictness of welfare rules. Factor 21 captures prior work experience interacted with child age.[27] Factor 24 captures CCDF spending, CSE, and EITC. Factor 26 captures benefit levels, remaining eligibility, EITC, and local wages.

Table 7 examines how well the 14 factor instruments predict cumulative child care. Their incremental R^2 in a first-stage regression is .0546 (vs. .0908 for all 78 instruments), and the F-test of their joint significance is 14.57 (p = .0000). The factors also have sensible coefficients in the reduced form. For instance, factor 6 has a substantial negative coefficient ($t = -8.8$), suggesting that higher benefit levels reduce work and child care use, as expected. Factor 12 also has a substantial negative coefficient ($t = -5.5$), implying that higher unemployment reduces work and child care. Factor 21 has a positive coefficient ($t = 7.6$), so work experience/cumulative child care grows more quickly with child age for mothers with more prior experience. (Note that factors with the most explanatory power for child care do not correspond to those that had the largest eigenvalues in the factor analysis.)

The Shea (squared) partial correlations for child care, income, and children are .0967, .0610 and .2222, respectively. Thus, the 14 factor instruments successfully generate independent variation of the three endogenous variables. For income, the most important factors are 1, 3, 7, 10, 24, and 26, while for number of children, they are 1, 3, 6, 8, 12, and 23. Factors that are particularly important for explaining income (but not child care) are 7 and 10, which capture work requirement exemptions and whether the woman could have hit a work requirement. For children, factor 23, which captures CSE, is particularly important. Factors 1 and 3 are im-

[27] Prior work experience and child age are included in the main equation, but their interaction is not. Recall that prior work experience is a proxy for the mother's skill endowment, which is correlated with the child's skill endowment. Thus, excluding the age interaction from the main equation follows logically from the assumption that the child skill endowment has an age-invariant effect in the test score equation (see the discussion prior to eq. [3]). The interaction is useful for predicting cumulative child care since mothers who work more prior to childbirth tend to work more afterward.

portant for both income and children (but not child care). Factor 1 captures several aspects of time limits and work requirements, while factor 3 captures whether a woman could have hit a work requirement, along with a range of exemptions.

Table 8 reports results for LIML, Fuller, 2SLS, and GMM estimators, using the 14 factors as instruments, and compares these to LIML using all 78 instruments. Three aspects of the results are notable. First, the LIML estimate of the effect of cumulative child care is hardly affected by using the 14 factors in place of the full set of 78 instruments. Second, and most important, using the factors as instruments leads to an increase in efficiency. The standard error of the LIML estimate of the child care coefficient is reduced by about 10%, and the t-statistic increases from -1.86 to -2.13.

Third, the LIML and 2SLS estimates are very similar when the reduced set of 14 instruments is used. The 2SLS estimate is now $-.00498$, with a t-statistic of -2.08. Thus, the 2SLS estimate is now shifted only 5% of the way toward OLS, versus 25% when using all 78 instruments. Indeed, the Cragg-Donald weak-instruments test statistic, which was 5.80 when we used all 78 instruments, improves to 15.33 using the 14 factors. In the case of three endogenous variables and 14 excluded instruments, the Stock-Yogo critical values for the null that the 2SLS bias may exceed 20%, 10%, or 5% of the OLS bias are 5.93, 10.25, and 18.47, respectively. Thus, using the 14 factors as instruments, we can now clearly reject the null that the 2SLS asymptotic bias may exceed 10% of the OLS bias. Using factor analysis to reduce the size of the instrument set is an effective way to reduce the 2SLS bias.

Table 8, column 6, reports the results from using the first 21 principal factors as instruments. As we saw in table 7, these 21 factors have less explanatory power for the endogenous variables than our 14 factors. Hence, we view this approach as inferior to ours. (See app. H in the online version of this article for further discussion.) However, it is comforting that this approach leads to similar results. That is, it gives just a slightly larger estimate of the child care effect (-0.63% per quarter, $t = -2.3$).

D. Robustness of the Results with Respect to Specification of the Main Equation

In table 9 we consider the sensitivity of our results to six changes in the specification of the main equation. First, we include additional controls for the mother's age at childbirth (age and age squared in addition to the under age 20 and over age 32 dummies). This has a negligible impact on the estimates—compare columns 1 and 2—as does eliminating age controls entirely (not reported).

What motivates this experiment is concern that our welfare policy in-

Table 8
Results Based on Factor Analysis of the Instruments

		14 Rotated Factors				
	LIML 78 Instruments (1)	LIML (2)	FULLER[a] (3)	2SLS (4)	GMM (5)	LIML Principal Factors (6)
Cumulative child care	−.00522[+]	−.00533*	−.00531*	−.00498*	−.00420[+]	−.00633*
	(.0028)	(.0025)	(.0025)	(.0024)	(.0023)	(.0027)
Log(cumulative income)	.01037	.01062	.01060	.01021	.01061	−.00094
	(.0242)	(.0266)	(.0265)	(.0243)	(.0213)	(.0269)
Mother's education	.01276*	.01297*	.01297*	.01282*	.01308*	.01298*
	(.0032)	(.0030)	(.0030)	(.0030)	(.0029)	(.0033)
Mother's AFQT	.00066*	.00066*	.00066*	.00065*	.00063*	.00077*
	(.0003)	(.0003)	(.0003)	(.0003)	(.0003)	(.0003)
R^2	.3847	.3844	.3845	.3860	.3890	.3761
k^{b}	1.040	1.005	1.005			1.006
Weak/many-instruments test	5.80	15.33	15.33	15.33	15.33	11.13

NOTE.—The dependent variable is log (test score). No. observations = 3,787. In col. 1, we use as instruments the same 78 variables described in the note of table 5. Additional control variables are listed in table 2. Instruments used in cols. 2–5 are 14 factors derived from the factor analysis of our original 78 instruments described in the note in table 5. Instruments used in col. 6 are the first 21 (unrotated) factors. Robust standard errors (Huber-White) by child clusters are in parentheses.

[a] Fuller parameter $\alpha = 1$.

[b] k is the parameter of the k-class estimator, which equals one for 2SLS and exceeds one for LIML.

[+] Significant at the 10% level.

* Significant at the 5% level.

Table 9
Robustness with Respect to the Specification of the Main Equation

	Baseline (1)	Additional Mother's Age Controls (2)	Removing AFQT (3)	Children by Ages (4)	Year Effects (5)	State Effects[a] (6)	Test Score in Levels[b] (7)
Cumulative child care	−.00533*	−.00535*	−.00613*	−.00553*	−.00490$^+$	−.00982$^+$	−.50258*
	(.0025)	(.0026)	(.0031)	(.0026)	(.0026)	(.0053)	(.2118)
Log(cumulative income)	.01062	.00259	.07808*	.00564	.00228	.02863	1.20063
	(.0266)	(.0279)	(.0284)	(.0280)	(.0300)	(.0344)	(2.2509)
Mother's education	.01297*	.01435*	.01355*	.01538*	.01329*	.01467*	1.05541*
	(.0033)	(.0036)	(.0044)	(.0031)	(.0036)	(.2532)	(.0030)
Mother's AFQT score	.00066*	.00070*		.00067*	.00067*	.00069*	.07814*
	(.0003)	(.0003)		(.0003)	(.0003)	(.0003)	(.0271)
Child's age	.03817*	.04054*	.01833	.03812*	.04078*	.04225*	3.38379*
	(.0122)	(.0126)	(.0128)	(.0124)	(.0131)	(.0131)	(1.0073)
Mother's age		−.01260	−.00101				
		(.0150)	(.0169)				
(Mother's age)^2		.00024	−.00005				
		(.0003)	(.0004)				
I[age of mother$_i$ < 20]	.02368*	.00799	.00811	.04335	.0089	.02387$^+$	1.83958$^+$
	(.0116)	(.0149)	(.0157)	(.0272)	(.0115)	(.0119)	(.9577)
I[age of mother$_i$ ≥ 33]	.0060	−.0049	.0032	.0109	.0003	.0081	.2064
	(.0256)	(.0310)	(.0347)	(.0269)	(.0257)	(.0261)	(2.2426)
Mother's age at first birth		−.0017					
		(.0019)					

No. children	−.02545*	−.02726*	−.02616*		−.0250*	−.02975*	−2.15419*
	(.0064)	(.0069)	(.0079)		(.0067)	(.0080)	(.5449)
No. children ages 0–5				−.0298*			
				(.0079)			
No. children ages 6–17				−.0095			
				(.0205)			
Year (at time of test)					−.0102*		
					(.0039)		
(Year)^2					.0006*		
					(.0003)		
R	.3844	.3842	.3304	.3808	.3884	.3646	.3936
k[c]	1.005	1.006	1.013	1.005	1.008	1.004	1.005
Weak/many-instruments test	15.33	13.04	19.85	11.85	12.99	9.02	15.33

Note.—The dependent variable is log (test score). N = number of observations = 3,787. The estimation method for all models is LIML. Instruments are 14 factors derived from the factor analysis of our original 78 instruments described in the note to table 5. Robust standard errors (Huber-White) by child clusters are in parentheses.

[a] Joint-significance test for state fixed effects = 30.25 (.066).

[b] The mean score is 91.9, so the point estimate implies a child care effect of −.55 per quarter, or −2.2% per year, similar to the log results.

[c] k is the parameter of the k-class estimator, which equals one for 2SLS and exceeds one for LIML.

[+] Significant at the 10% level.

[*] Significant at the 5% level.

struments are correlated with mothers' age at childbirth due to the timing of waivers/TANF. Waivers were first implemented in some states in 1992–93 and, as Fang and Keane (2004) note, binding work requirements first hit significant numbers of women in 1995–96. Thus, for a child born prior to 1990, it is unlikely that waivers could have influenced the mother's labor supply before the child was age 6. In the NLSY79, women who had children prior to 1990 tended to be younger at childbirth than those who had children later. Indeed, from 1990 onward, all births are to mothers in their twenties and thirties, while prior to 1990, many were to teenage mothers. So, loosely speaking, stricter welfare rules and greater child care usage will be positively correlated with maternal age at childbirth. Then, (i) if mother's age at childbirth has a positive effect on child cognitive ability and (ii) if we fail to adequately control for mother's age at childbirth in the main equation, this will generate a spurious positive effect of maternal work/child care use on child cognitive ability test scores.

Results in columns 1 and 2 of table 9 suggest that this is not a concern. Conditional on measures of mothers' human capital and economic resources (i.e., education, AFQT), maternal age at childbirth is not positively correlated with children's achievement. Indeed, if anything, the results suggest that maternal youth is beneficial for child outcomes (the under age 20 dummy is positive in col. 1). Column 2 includes additional controls for mother's age at childbirth (for the present child and the first child), but they are not significant. (Appendix I in the online version of this article contains more evidence on this issue.)

Next, in table 9, column 3, we consider the impact of dropping the mother's AFQT score. This leads to a slight increase in the child care coefficient (from −2.1% to −2.4% per year). More noticeable is that it produces a large increase in the cumulative income coefficient, which becomes highly significant. This is consistent with the view that lifetime income is more important than transitory income in determining parental investment in children and, hence, children's achievement: with AFQT omitted, income is significant, as it proxies for the mother's permanent income/skill endowment; however, AFQT is a better proxy, so when it is included, the income variable drops out.[28]

Third, we consider the sensitivity of our results to controls for the ages of siblings. Our baseline model controls for the number of siblings but not their ages. If younger children have a different effect on the mother's time constraint, this may bias our estimated child care effect. Thus, column

[28] Even without AFQT, the implied effect of income remains modest. The point estimate implies that, at the mean of the data, a doubling of cumulative income increases test scores by about (.078)(.69) = 5.4%. However, the model still includes such variables as mother's education and prechildbirth wage, which are also proxies for her permanent income/skill endowment.

4 includes separate controls for number of children aged 0–5 and 6–17 (both treated as endogenous). The estimates imply that young siblings (ages 0–5) have a larger negative effect on test scores. However, this has little impact on the estimated child care effect, which is now −2.2% per year.

Fourth, we consider aggregate time effects. It is possible that during our sample period an omitted time-varying factor both influenced child test scores and was correlated with the increasing stringency of welfare rules. Column 5 includes a quadratic time trend to address this concern. Interestingly, the quadratic has a U-shape, implying that an aggregate factor not included in our model first drove down test scores followed by a recovery. But including the time trend only slightly reduces the estimated child care effect, from −2.1% to −2.0% per year. Thus, any bias from omitted time effects appears to be minor. Results were essentially identical using unrestricted time dummies.

Next, in column 6 we add state fixed effects. Many researchers would prefer a fixed effects specification since it deals with any (cross-state) correlation between the instruments and unobserved child ability endowments (e.g., if states with relatively low-ability children adopted stricter welfare reform, it would bias the child care effect negatively). Adding state fixed effects to the main equation actually shifts our estimate of the cumulative child care effect from −.53% to a much larger negative value of −.98%. However, it also reduces the precision of the estimate, more than doubling the standard error. Despite this, the estimate remains significant at the 10% level ($t = -1.85$).

But we are skeptical of fixed effects for several reasons. On a priori grounds, we are skeptical that child ability differs systematically by state (conditional on extensive controls, such as mother's AFQT and education). Consistent with this, state fixed effects are not significant at the 5% level (see note a of table 9), and state effects explain only 2% of the variance of the residuals in our baseline model. Even if unobserved child ability differed by state, it would only induce bias if it were correlated with the instruments, for example, if states with low-ability children adopted stricter welfare rules. But in appendix J in the online version of this article we show that there is no significant difference in average child test scores (in the prereform period) between states that adopted more versus less strict reforms.

Furthermore, the fixed effects estimator should be used with caution. It seems implausible that the strict exogeneity assumption required for its consistency would hold in the child production function context.[29]

[29] The strict exogeneity assumption fails if child test score realizations at age t affect future inputs into child cognitive ability production and/or how the welfare policy rules evolve. (See also the discussion of this point in Sec. II.)

Keane and Wolpin (2002) show that fixed effects can give very misleading results if average and deviations from average values of policy variables have different effects on current decisions.[30] Given these considerations, we do not adopt fixed effects as the baseline model.

Finally, in table 9, column 7, we consider a model where test scores enter in levels. The estimates imply that one quarter of child care reduces scores by 0.5 points. As the mean score is 91.9, this corresponds to −0.54% per quarter, which is almost identical to the effect in the log specification.

E. Robustness of the Results with Respect to the Instruments

It is well known that IV estimates can be sensitive to the instrument list and that, given unobserved heterogeneity in treatment effects, what IV identifies depends on the instruments used (see Heckman and Vytlacil [2005] for an extensive discussion of this issue). Thus, it is important to examine the robustness of our results to the instrument list. Table 10 reports LIML results using the baseline list of 78 instruments in column 1 and seven variants on that list in columns 2–8. To experiment with dropping certain types of instruments, we must use the full set of 78 instruments. Table 11 examines the explanatory power of the different instrument sets in the first-stage regressions for child care.

In column 2, of table 10, we exclude CCDF spending. It arguably belongs in equation (6) since it shifts the effective price of child care, but excluding this instrument has little impact on the estimated child care effect. In column 3 we use only the main features of TANF as instruments: time limits, work requirements, and disregards. This increases the estimated child care effect to −3.0% per year. Column 4, in contrast, drops TANF-related instruments, using other aspects of the policy/demand environment to identify the child care effect. This reduces the estimate slightly to −1.7% per year.

In column 5 we drop all instruments specific to the welfare reforms of the 1990s (e.g., TANF, CCDF, and EITC), using only instruments that varied across states/time regardless. These are state welfare grant levels and local demand conditions. Here the child care effect estimate is only slightly smaller than in our baseline, −1.75%, and it is significant at the 10% level ($t = -1.82$).

In our reduced form regression, we interact all policy and demand variables with mother's education and AFQT. This lets changes in policy/

[30] A state fixed effect controls for a state's average level of welfare generosity. Thus, using state effects, we estimate the impact of deviations from the average level of child care use induced by deviations from average welfare rules. Such short-run effects may differ from the effects of long-run policy changes, and the latter are presumably of greater interest.

Table 10
Robustness with Respect to the Instrument List

	Original Set of IVs[a] (1)	Excluding CCDF[b] (2)	Only TL, WR, and ED (3)	Excludes TL, WR, and ED (4)	Only BEN and Local Demand (5)	Original Set without Interactions (6)	Only State-Specific Instruments[c] (7)	Only State-Specific IVs[c] without Interactions (8)
Cumulative child care	−.00522$^+$	−.00584*	−.00758*	−.00414$^+$	−.00437$^+$	−.00528$^+$	−.00623*	−.00642*
	(.0028)	(.0028)	(.0038)	(.0024)	(.0024)	(.0028)	(.0025)	(.0026)
Log(cumulative income)	.01037	.01026	.02118	.02054	−.00294	−.00445	−.00300	.00529
	(.0242)	(.0250)	(.0367)	(.0302)	(.0372)	(.0239)	(.0246)	(.0361)
Mother's education	.01276*	.01305*	.01327*	.01049*	.01208*	.01285*	.01316*	.01167*
	(.0032)	(.0032)	(.0034)	(.0036)	(.0036)	(.0030)	(.0034)	(.0037)
Mother's AFQT	.00066*	.00067*	.00064*	.00056$^+$	.00073$^+$	.00077*	.00078*	.00073$^+$
	(.0003)	(.0003)	(.0003)	(.0003)	(.0004)	(.0003)	(.0003)	(.0004)
R^2	.3847	.3816	.3705	.3843	.3860	.3820	.3767	.3720
k^d	1.040	1.039	1.036	1.004	1.003	1.008	1.026	1.006
Weak/many–instruments test	5.80	5.80	5.21	7.00	5.71	9.09	6.27	7.00
P-value, Hansen J-statistic	.744	.713	.458	.979	.826	.892	.787	.911
No. instruments	78	75	58	27	18	26	63	25

Note.—The dependent variable is log(test score). The estimation method is LIML. Robust standard errors (Huber-White) by child clusters are in parentheses.

[a] All 78 policy variables, local demand conditions, and interactions described in note to table 5. Unless noted in the column heading, all specifications still include these interactions.

[b] See descriptions of instruments in table 1. CCDF = Child Care Development Fund expenditures; TL = time limits; WR = work requirements; ED = earnings disregards; BEN = benefit amounts.

[c] Excludes all individual-specific welfare rules, such as, whether a woman could have hit a time limit or a work requirement (i.e., all variables with an i subscript in table 1).

[d] k is the parameter of the k-class estimator, which equals one for 2SLS and exceeds one for LIML.

$^+$ Significant at the 10% level.

* Significant at the 5% level.

Table 11
Explanatory Power of Instruments in First-Stage Regressions for Child Care (Instruments in Table 10)

Instruments Used in Each Column in Table 10	(Partial Correlation)^2	(Shea Partial Correlation)^2	Incremental R^2	F-Statistic	p-Value
Original set of IVs	.1735	.1483	.0908	14.740	.000
Excluding CCDF	.1675	.1426	.0877	14.980	.000
Only TL, WR, and ED	.1307	.0944	.0684	18.120	.000
Excludes TL, WR, and ED	.1151	.1094	.0603	11.800	.000
Only BEN and local demand	.1046	.0985	.0548	16.210	.000
Original set without interactions	.1327	.1072	.0755	11.930	.000
Only state-specific IVs	.1443	.1393	.0756	9.550	.000
State-specific IVs without interactions	.1189	.1144	.0622	15.510	.000

The dependent variable is cumulative child care. R^2 of the first-stage regression with only exogenous variables = .4765.

demand have different effects on different types of mothers (e.g., welfare rules are less important for the college educated). In column 6 we drop these interactions to see how important they are. This has little impact on the estimated child care effect.

Recall that some of our instruments are tailored to individuals based on ages of their children (see Sec. III.A). In column 7 we drop these individual-level instruments. The resulting estimate is −2.5% per year, which is slightly larger than our baseline estimate. In column 8 we go further and also drop interactions of the instruments with mother's education and AFQT. Thus, we rely purely on state-level variation to identify the child care effect. This gives an estimate of −2.6% per year.

The use of state-level instruments begs the question whether we should cluster standard errors at the state level (instead of the child level). One might also consider clustering at the mother level, as 368 out of the 944 mothers in our sample have multiple children. But in appendix K in the online version of this article, we show that this makes little difference. Within-cluster correlations are quite small at these levels. Clustering by child increases standard errors by 25%–40%, depending on the instrument set. Clustering by mother or state has a very similar effect.

In summary, our result of a negative child care effect is robust to a wide range of alternative instrument sets, with estimates ranging from −1.7% to −3.0% per year and with all but one estimate between −1.7% and −2.5% (compared to our baseline of −2.1%). We experimented with many other instrument sets (not reported) and continue to find results in this general range. Finally, the Hansen J-test does not reject the overidentifying restrictions for any instrument set we consider (see the next to the last row of table 10).

F. Heterogeneity in the Effect of Child Care by Type of Mother and Child

In table 12, we assess how the effects of child care vary with characteristics of the mother or child. In columns 1–3 we include interactions between cumulative child care and mother's education, AFQT, and number of children. These variables are de-meaned before being interacted. Thus, main effect estimates can be interpreted as the effect of child care for a typical mother.

In table 12, column 1, the interaction between mother's education and child care is negative, as expected (i.e., time of less educated mothers is less valuable for child ability production). Its t-statistic is -1.76, so it is only significant at the 8% level, but the point estimate is fairly substantial. It implies, for example, that if a mother's education is 4 years above the sample average, then the negative child care effect goes from $-.46\%$ to $-.81\%$. The later estimate has a standard error of .28 and hence a t-statistic of -2.90. Thus, we have clear evidence that child care has a more negative effect if the mother is more educated. In column 2, we see the same pattern for AFQT. However, in column 3, the interaction between cumulative child care and number of children in the household is very small and insignificant, implying that the effect of child care does not depend on number of siblings.

Column 4 includes an interaction of child care with gender. It is significant, and the point estimates imply that the effect of child care is -0.71% per quarter for girls but only -0.38% for boys. Column 5 includes an interaction with race (nonwhite = 1). It is not significant, although the point estimates imply a larger negative effect for whites. A similar result is obtained in column 6, where we include separate black and Hispanic dummies. Thus, we find little evidence of race differences. Finally, column 7 allows the effect of child care to vary linearly with age.[31] The age/child care interaction variable is insignificant, so we find no evidence of differences by child age.

G. The Effects of Different Types of Child Care

So far we have reported on effects of child care in general, but it seems likely that the type of care matters. That is, formal center-based care by

[31] We let the impact of child care be a linear function of child age, as in $\sum_{\tau=1,t}(\phi_{20}+\phi_{21}t)C_t$. This expression can be rewritten $\phi_{20}C_t+\phi_{21}\sum_{\tau=1,t}(tC_t)$, where $\sum_{\tau=1,t}(tC_t)$ is the age/child care interaction variable referred to in the text. We add this variable to eq. (9) and treat it as endogenous. The incremental R^2 for the excluded instruments in the reduced form regression for this variable is .103, with a F-statistic of 25.3 (see table 5, panel B).

Table 12
Heterogeneity in the Effect of Maternal Time Inputs

	By Maternal Education (1)	By Maternal AFQT Score (2)	By Number of Children (3)	By Child's Gender (4)	By Child's Race (5)	By Child's Race[a] (6)	By Child's Age (7)
Cumulative child care	−.00456	−.00493+	−.00502+	−.00709*	−.00814*	−.00885+	.00465
	(.0029)	(.0029)	(.0027)	(.0029)	(.0041)	(.0046)	(.0145)
$\widetilde{\text{Education}}$ × (Cumulative child care)	−.00088+						
	(.0005)						
$\widetilde{\text{AFQT}}$ × (Cumulative child care)		−.00011+					
		(.00007)					
$(\widetilde{\text{Number of children}})$ × (Cumulative child care)			−.00008				
			(.0013)				
Male × (Cumulative child care)				.00331*			
				(.0015)			
Nonwhite × (Cumulative child care)					.00371		
					(.0030)		
Black × (Cumulative child care)						.00469	
						(.0031)	
Hispanic × (Cumulative child care)						.00325	
						(.0036)	
Age-weighted cumulative child care[b]							−.00104
							(.0015)

Log(cumulative income)	.01379	.01441	.01004	.00661	.00827	.00740	.01089
	(.0244)	(.0258)	(.0244)	(.0243)	(.0245)	(.0246)	(.0242)
Mother's education	.01858*	.01232*	.01288*	.01313*	.01278*	.01276*	.01251*
	(.0042)	(.0032)	(.0031)	(.0032)	(.0031)	(.0031)	(.0032)
Mother's AFQT score	.00066*	.00171*	.00065*	.00068*	.00072*	.00076*	.00067*
	(.0003)	(.0007)	(.0003)	(.0003)	(.0003)	(.0003)	(.0003)
R^2	.3868	.3874	.3861	.3828	.3867	.3883	.3831
k^c	1.0390	1.0412	1.0400	1.0403	1.0405	1.0413	1.0396
Weak/many–instruments test	5.73	5.72	5.61	5.73	5.72	5.63	5.00
Test for joint significance of interactions						2.47	
						(.2905)	

NOTE.—The dependent variable is log(score). Estimation method is LIML. N = number of observations = 3,787. $\widetilde{\text{Education}}$ = Education − $\overline{\text{Education}}$, where $\overline{\text{Education}}$ is the mean (same for number of children and mother's AFQT score). Instruments are the same 78 variables described in the note of table 5 plus predicted child care from the first stage interacted with mother's education (col. 1), with mother's AFQT (col. 2), with number of children (col. 3), with gender dummy (col. 4), with nonwhite dummy (col. 5), and with black dummy and Hispanic dummy (col. 6). Robust standard errors (Huber-White) by child clusters are in parentheses.

[a] The baseline includes AFQT, black, and Hispanic interacted with test dummies.

[b] Weights are the age of the child in number of quarters, as described in n. 31.

[c] k is the parameter of the k-class estimator, which equals one for 2SLS and exceeds one for LIML.

[+] Significant at the 10% level.

* Significant at the 5% level.

trained providers (e.g., day care centers, preschool) may have different effects from informal care provided by relatives (e.g., grandparents, siblings) or nonrelatives. Thus, we estimated versions of equation (9) in which the effects of child care are allowed to vary by type of care.[32] Table 5, panel B, presents the results of the reduced form regressions for child care inputs of different types (formal, informal, etc.). Even at this more refined level, the welfare policy/demand condition variables are reasonably powerful instruments. The marginal R^2s for the excluded instruments in the reduced form regressions range from .086 to .099, Shea partial correlations (squared) range from .10 to .15, and joint F-tests show that the excluded instruments are highly significant.

The reduced form regression coefficients (not reported) also appear to be reasonable. In general, women are more likely to use informal care relative to formal care in states with stricter rules. Specifically, mothers are more likely to use formal versus informal care if (i) a state does not have a work requirement, (ii) it has young child or other work requirement exemptions, (iii) it has a longer work requirement time limit, (iv) work requirements were implemented more recently, (v) less time has elapsed since a time limit could have hit, (vi) remaining eligibility is greater, (vii) a state has higher CCDF spending, or (viii) earnings disregards are greater. If a state has more exemptions, it reduces the probability of using child care in general, but that of using informal care is reduced much more. Education interactions imply that welfare rules have less influence on more educated women.[33]

Strikingly, the LIML results in table 13 indicate that formal (i.e., center-based) care does not have any adverse effect on cognitive outcomes. Only informal care leads to significant reductions in achievement. In particular, an additional year of informal child care causes a 2.6% reduction in test scores. The estimated effect for formal care is actually positive, but it is insignificant.

Our finding that informal care has adverse affects relative to formal care is arguably the most important of this study in that it may provide a rationale for government programs (like CCDF in the United States or

[32] Recall that we do not have direct measures of child care use in years 4 and 5, and we impute this using the procedure described in Sec. V.B and in app. B. Having imputed child care use, we now impute whether it was formal or informal by looking at the last available observation on type of care used. This should not induce much error because the degree of persistence in type of care is tremendous. Conditional on using child care for two consecutive periods, the own transition rates for formal and informal care are both roughly 98%.

[33] The AFQT effects are subtler, but they are also sensible. A high AFQT reinforces the effect that disregards and exemptions increase chances of using formal care. The chance of using formal care increases with the time elapsed since a state imposed time limits, but it does so only for relatively high-AFQT women (consistent with high-AFQT women being more likely to find work).

Table 13
Child Care Effects by Type of Care

		Formal versus Informal Child Care		Care Provided by Relatives versus Nonrelatives		Formal/Informal by Education
	Mean (1)	OLS (2)	LIML[a] (3)	OLS (4)	LIML (5)	LIML (6)
Cumulative informal child care	5.8533	.00067	−.00643*			−.00638*
	(5.873)	(.0008)	(.0029)			(.0031)
Relatives	5.0077			.00034	−.00751*	
	(5.736)			(.0008)	(.0029)	
Nonrelatives	1.1454			.00164	.00683	
	(3.355)			(.0011)	(.0069)	
Informal child care × I[mother college graduate]						−.00032
						(.0023)
Cumulative formal child care (day care, nursery, prekindergarten, other)	1.2229	.00307*	.00302	.00309*	.00472	.00831
	(3.055)	(.0011)	(.0066)	(.0011)	(.0066)	(.0129)
Formal child care × I[mother college graduate]						−.00703
						(.0101)
Log(cumulative income)	3.6332	−.00343	.00719	−.00331	.00883	.01094
	(.730)	(.0057)	(.0233)	(.0057)	(.0230)	(.0233)
R^2		.4004	.3776	.4007	.3520	.3722
k[b]			1.0385		1.0347	1.0379
Weak/many-instruments test			4.5		4.48	3.3

Note.—Dependent variable is log (score). N = number of observations = 3,787. Instruments are the same 78 variables described in the note to table 5 plus predicted formal and informal child care from the first stage interacted with a dummy variable for whether the mother is a college graduate or not (only in the last column). Robust standard errors (Huber-White) by child clusters are in parentheses.

[a] If state fixed effects are included in this regression they are not significant at the 10% level (see online app. I, table I1). The coefficient on formal care remains small and insignificant, while that on informal care increases in magnitude to −.01013 (SE = .0061).

[b] k is the parameter of the k-class estimator, which equals one for 2SLS and exceeds one for LIML.

* Significant at the 5% level.

Child Care Benefit in Australia) that create incentives for mothers to use formal rather than informal care. Thus, we subjected this result to the same battery of robustness tests we applied to our estimates of the effect of child care in general. The results are reported in appendix I in the online version of this article, and they show that the result is quite robust to changes in specification, the instruments, and so forth.[34]

In table 13, columns 4 and 5, we divide informal child care into that provided by relatives (most often grandparents) versus nonrelatives (e.g., family day care). Here we find that only informal care by relatives has a negative effect. Note that informal care by relatives is the most common arrangement for the single mothers in our sample (60%). Informal care by nonrelatives accounts for a little less than 20%, and formal center-based care accounts for a bit over 20%. This preponderance of informal care explains why our overall estimate of the effect of child care is negative (i.e., −2.1% per year).

Our results here are basically consistent with work by Hansen and Hawkes (2008). Looking at Bracken school readiness scores in the Millennium Cohort Survey, they find a negative effect of grandmother care relative to formal center-based care. Similarly, Gregg et al. (2005), using the Avon Longitudinal Survey, find that early maternal employment reduces subsequent child test scores only if children were placed in informal care (i.e., care by a relative or friend).

It may seem surprising that care by relatives—predominately grandparents—leads to worse outcomes since grandparents presumably care a great deal about grandchildren. But there is a literature in sociology showing that grandparents often find caring for young children stressful and physically demanding (Millward 1998; Goodfellow and Laverty 2003). Prior literature also suggests that center-based care has two advantages over informal care: (*a*) trained care providers may provide more cognitive simulation to children than informal providers,[35] and (*b*) center-based care may provide more stimulating interaction with other children and more educational activity than informal care.[36]

[34] In particular, with state fixed effects, the coefficient on formal care remains small and insignificant, while that on informal care is still large (−.01013) and significant ($t = -1.66$). However, state effects are not significant at the 10% level.

[35] McCartney (1984), Melhuish et al. (1992), and NICHD (2000) find a key difference between high- and low-quality care is the amount of language stimulation. Center-based teachers are more likely to have training in child development and to be more educated in general, both of which are associated with more verbal stimulation. According to NICHD (2000), they also tend to provide more supportive, attentive, and interactive care.

[36] Ideally, we would also like to examine how the effects of child care differ by direct measures of child care quality. However, the NLSY lacks good quality measures. Hence, we have instead differentiated between formal and informal care. However, the notion that formal care is superior is consistent with the

In column 6 of table 13, we interact type of child care with a dummy for whether the mother has some college education. The point estimates imply a substantial positive effect of formal care for low-education mothers (i.e., +0.83% per quarter), but it is very imprecisely estimated. Still, this result is at least not inconsistent with prior results suggesting that center-based care is beneficial for low socioeconomic status (SES) children (see, e.g., Currie and Thomas [2005] on Head Start or Pungello, Campbell, and Barnett [2006] on the Perry Preschool and Abecedarian experiments).

H. The Effect of Welfare Reform on Child Test Scores

Our focus has been on using welfare policy changes of the mid-1990s as a source of exogenous variation to help identify effects of child care on child outcomes. But it is also interesting to examine the effect of welfare reform itself. We do this using the reduced form for child test scores, obtained by substituting the excluded instruments in table 1 for the three endogenous variables in equation (9). One change is that we include a quadratic time trend. While we found this made little difference to the estimated child care effect, we believed that it was important to control for possibly omitted time effects in the reduced form to avoid the risk of attributing the impact of any such factors to welfare reform.

The 78 instruments are highly significant in the reduced form test score equation–the *F*-test for their joint significance is 2.79 compared to the 1% critical value of 1.44. In a simpler model that leaves out interactions with mother's education and AFQT, the remaining 30 instruments give an *F*-test of 2.20, compared to the 1% critical value of 1.70. These results suggest that changes in welfare rules did have a significant impact on child test scores. However, given the complexity of the set of variables that characterize the welfare rules, it is quite difficult to put a meaningful interpretation on the individual coefficients. So we instead use the estimated reduced form to simulate the effect of changes of welfare rules on test scores. We simulate (and compare) average test scores under two scenarios: (i) using the policy variables that were actually in place and (ii) holding the policy variables fixed at a baseline level. (Fang and Keane [2004] used a similar procedure to evaluate effects of changes in welfare rules on employment and welfare participation.)

---

evidence on who uses it. In online app. M we present a logit for whether a mother uses formal or informal care (conditional on child care use). The results show that more educated, urban women with fewer children are more likely to use formal care. This suggests that formal care is higher quality since it is typically used by women who can afford more expensive care. Similarly, online app. M also presents a logit for use of relatives vs. nonrelatives (conditional on using informal care). The more educated, urban women with fewer children are more likely to use nonrelatives, suggesting that nonrelatives provide higher quality care than relatives.

Simulation of the reduced form model implies that changes in welfare rules had almost no impact on child test scores during the 1979–93 period. This is not surprising as the rule changes during that period were modest. However, the rule changes began to reduce test scores after 1993. Our model implies that average test scores for children of single mothers in the 1994–99 period were 1.32% lower than they would have been had the rules not changed. This figure is broadly consistent with our point estimate for the effect of child care. As we saw in table 4, the child care usage rate was about 10 percentage points higher in the post-93 period. Thus, by 1999 (6 years later) children would have had about 0.6 extra years of child care on average. Multiplying this by our estimated annual effect of −2.13%, we obtain −1.3%, which is quite close to our simulated effect of −1.32%.

I. Child Care and Noncognitive Outcomes

Of course, child care may also affect noncognitive outcomes. Indeed, it is possible that any negative effects on cognitive outcomes could be outweighed by positive effects on noncognitive outcomes. Table 14 reports a preliminary analysis of this issue, using the CNLSY79's behavioral problems index (BPI). The BPI measures incidence of 28 types of problem behaviors (i.e., antisocial behavior, hyperactive behavior, and depressed/withdrawn behavior) among the surveyed children.

The OLS point estimate implies essentially no effect of child care on the incidence of behavioral problems. However, the LIML results suggest a quantitatively large adverse effect of +0.46% per quarter of child care.[37] However, the t-statistic on this estimate is only 1.44, the low level of significance arising in part because the sample size for this regression ($N = 1{,}730$) is less than half that used in the cognitive ability regressions. Still, the result provides evidence against any claim that positive effects on behavioral outcomes might outweigh the negative effects on cognitive outcomes. More work is needed to look at other dimensions of noncognitive skill.

VII. Conclusions

In this article we have used the children of single mothers in the NLSY79 to assess the impact of child care use on child cognitive achievement measured at ages 3–6. To deal with endogeneity of child care, we utilize the (plausibly) exogenous variation in work/child care choices of

[37] We use 19 factors, which are chosen by the same procedure described in Sec. VI.C. Because the sample here is different ($N = 1{,}730$), the factor analysis results differed slightly, as did the regressions of the endogenous variables on the factors. We retained more factors because we used a lower significance cutoff ($t = 2$), as seemed appropriate given the smaller sample size.

Table 14
Effect of Child Care on Noncognitive Outcomes

	OLS	LIML[a]
Cumulative child care	−.00083	.00460
	(.0008)	(.0032)
Log(cumulative income)	−.00820	.03270
	(.0057)	(.0359)
No. of children	−.00448	.01385*
	(.0034)	(.0070)
I[mother's age<20]	−.01138	−.00523
	(.0099)	(.0117)
I[mother's age>30]	−.00508	−.01058
	(.0284)	(.0334)
Mother's education	−.00444+	−.00587+
	(.0027)	(.0035)
Mother's AFQT score	−.00005	−.00045
	(.0003)	(.0004)
I[nonwhite]	−.03239*	−.04504*
	(.0119)	(.0136)
Male	.03583*	.03699*
	(.0075)	(.0080)
Birthweight	−.01446*	−.01596*
	(.0062)	(.0066)
I[work before]	−.00152	−.01579
	(.0103)	(.0149)
I[work before] × skill	−.00939	−.01950+
	(.0093)	(.0117)
Experience before childbirth	.00256	−.00427
	(.0056)	(.0077)
Experience × mother's age	−.00012	.00004
	(.0002)	(.0002)
I[never married]	−.09500*	−.08624+
	(.0301)	(.0503)
I[separated]	−.09747*	−.10221*
	(.0310)	(.0547)
I[divorced][b]	−.10255*	−.10005+
	(.0324)	(.0532)
I[urban]	.00214	.01572
	(.0108)	(.0130)
Child's age	−.00036	−.02148
	(.0069)	(.0134)
k[c]		1.018
Weak/many-instruments test		5.08

NOTE.—The dependent variable is log (Behavioral Problems Index). N = number of observations = 1,730. Robust standard errors (Huber-White) by child clusters are in parentheses.

[a] Instrument list: 19 factors derived from the factor analysis of our original 78 instruments described in the note to table 5. This factor analysis was run specifically for the sample for whom we observe the BPI test score.

[b] Excluded category is widowed.

[c] k is the parameter of the k-class estimator, which equals one for 2SLS and exceeds one for LIML.

+ Significant at the 10% level.

* Significant at the 5% level.

single mothers generated by differences in welfare rules across states and over time. Our approach is motivated by the fact that the 1996 welfare reform, as well as earlier state welfare waivers, generated substantial new incentives for single mothers to work and use child care. This event provides a good opportunity to extend the literature on the effects of child care on child outcomes—a literature that has been limited by the difficulty of finding plausible instruments for child care use.

Our main results indicate that the effect of child care on children's achievement is negative, significant, and rather sizable. Estimates of our baseline model imply that 1 year of full-time child care reduces cognitive ability test scores by roughly 2.1%. This corresponds to 0.114 standard deviations, so it is a substantial effect. This estimate is quite robust across a wide range of specifications and instrument sets.

But this general finding masks important differences across types of child care, types of children, and types of mothers. What drives the negative estimate of the child care effect is that most (i.e., about 75%) child care used by single mothers is informal (i.e., care by grandparents or other relatives or by nonrelatives in non-center-based settings). Our estimates imply that a year of informal child care reduces child test scores by 2.6%. In contrast, we find that formal center-based care has no adverse effect on child outcomes.

In addition, we find that child care has a larger adverse affect on cognitive outcomes for girls than for boys and for children of more educated mothers. The latter is not surprising, as education presumably increases the value of maternal time in child cognitive ability production. We do not find significant differences by child age or race/ethnicity.

Prior work has related test scores measured as early as age 7 to later life outcomes. We extend this by showing that scores at even earlier ages (i.e., ages 4–6) are significantly related to completed schooling. For example, we find that a 1% increase in PIAT math scores at age 6, holding parental background variables such as mother's education fixed, is associated with an increase in educational attainment (measured at age 18 or later) of approximately .019 years. For reading scores, the figure is .025 years. Thus, for example, a 2.6% reduction in test scores induced by a year of full-time informal child care translates into roughly a .050 to .065 year reduction in completed schooling.

Appendix A

Table A1
Effect of Early Cognitive Ability Test Scores on Highest Grade Completed by 2000

	PPVT at Age 4		PIAT Math at Age 5		PIAT Reading at Age 5		PIAT Math at Age 6		PIAT Reading at Age 6	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Test score	.00819*	.01574*	.00633	.01627*	.00960*	.02092*	.01908*	.03166*	.02493*	.0374*
	(.0041)	(.0035)	(.0046)	(.0044)	(.0048)	(.0045)	(.0049)	(.0045)	(.0056)	(.0055)
Age of completed education measure[a]	.16449	.06336	.69629*	.68394*	.69097*	.66007*	.45305*	.41675*	.45629*	.40288*
	(.1563)	(.1575)	(.0752)	(.0717)	(.0758)	(.0723)	(.0438)	(.0409)	(.0439)	(.0411)
Highest grade completed by mother	.09231*		.05216+		.04901		.09646*		.10179*	
	(.0403)		(.0348)		(.0343)		(.0270)		(.0268)	
Highest grade completed by father	.02147*		.02069*		.01948*		.00833		.01065+	
	(.0083)		(.0076)		(.0075)		(.0064)		(.0064)	
Number of siblings	−.14160*		−.14066+		−.12912*		−.08882+		−.08942*	
	(.0586)		(.0543)		(.0535)		(.0428)		(.0424)	
Birth order	−.11146		−.13111		−.09435		−.11223		−.07853	
	(.0979)		(.0957)		(.0946)		(.0754)		(.0751)	
Race (1 = nonwhite)	.06958		.08739		.06939		−.06182		−.21639+	
	(.1751)		(.1523)		(.1496)		(.1258)		(.1243)	
Gender (1 = male)	−.36024*		−.42114*		−.42716*		−.39478*		−.37505*	
	(.1380)		(.1236)		(.1228)		(.1011)		(.1008)	
Mother's age at child's birth	−.03878		−.01219		−.02523		.02586		.03390	
	(.0387)		(.0336)		(.0331)		(.0282)		(.0280)	
Mother's AFQT score	.00389		.00378		.00450		.00128		−.00030	
	(.0038)		(.0033)		(.0033)		(.0029)		(.0028)	
Constant	7.2531*	8.3078*	−2.8778	−3.8171*	−2.9770	−4.0097*	−.6925	−.1622	−1.5869	−.6049
	(3.1866)	(2.9599)	(1.8248)	(1.4088)	(1.7977)	(1.3892)	(1.2501)	(.8931)	(1.2644)	(.9295)
N	363	363	451	451	446	446	747	747	739	739
R^2	.1578	.0537	.2791	.2014	.2912	.2209	.2365	.1761	.2457	.1760

Note.—Dependent variable is highest grade completed by 2000. Sample consists of young adults 18 years or older. All regressions are estimated by ordinary least squares. PPVT = Peabody Picture Vocabulary Test; PIAT= Peabody Individual Achievement Test. N = number of observations. Standard errors are in parentheses.

[a] The age of the young adult by 2000 if she is older than 18 years old. The average age is 21.8.

+ Significant at the 10% level.

* Significant at the 5% level.

Table A2
Cognitive Ability Tests in Our NLSY Sample: Descriptive Statistics

	PPVT			PIAT-Math		PIAT-Reading	
Child's age	3	4	5	5	6	5	6
Log(test score) in our sample	4.367	4.2689	4.402	4.539	4.543	4.633	4.606
	(.191)	(.295)	(.239)	(.152)	(.128)	(.152)	(.095)
Test scores in our sample	80.263	74.334	83.767	94.719	94.802	104.089	100.585
	(14.952)	(19.512)	(17.504)	(14.329)	(11.727)	(15.319)	(9.462)
Nonwhites	78.007	70.836	82.135	93.836	94.247	103.358	100.482
	(14.169)	(17.958)	(16.889)	(14.289)	(11.685)	(15.454)	(9.269)
Whites	92.167	89.299	93.852	99.576	97.657	108.100	101.112
	(13.348)	(18.885)	(18.001)	(13.634)	(11.578)	(13.970)	(10.422)
Maternal education (12 years +)	82.820	78.748	88.743	97.084	96.823	106.755	102.265
	(14.369)	(18.917)	(17.648)	(14.178)	(11.663)	(15.131)	(9.425)
Maternal education (<12 years)	76.301	68.748	79.508	91.767	92.751	100.697	98.847
	(15.025)	(18.847)	(16.245)	(13.991)	(11.449)	(14.909)	(9.197)
Male	79.753	72.242	83.035	93.726	93.710	102.557	99.232
	(14.664)	(20.048)	(18.143)	(14.307)	(12.292)	(15.563)	(9.404)
Female	80.707	76.299	84.569	95.739	95.827	105.685	101.838
	(15.225)	(18.820)	(16.783)	(14.305)	(11.091)	(14.922)	(9.357)
N	339	512	438	598	663	584	653

NOTE.—PPVT = Peabody Picture Vocabulary Test; PIAT = Peabody Individual Achievement Test. *N* = number of observations. Standard deviations are in parentheses.

Appendix B

Table B1
Probit to Predict Child Care Choices of Nonworking Women in Years 4 and 5 after Childbirth

Variable	Coefficient
Whether worked before giving birth	.5920*
	(.208)
(Whether worked before) × (Average wage before)	−.0642†
	(.040)
Total work experience (prior to giving birth)	−.0060
	(.019)
Child's race	−.0874
	(.170)
Child's gender	.0497
	(.120)
Mother's education	.0821*
	(.038)
Total work experience since child birth	−.3983*
	(.070)
Total child care use since child birth	.2226*
	(.053)
Whether used child care or not in $t - 1$	1.7801*
	(.164)
Pseudo-R^2	.4585

NOTE.—Dependent variable is Pr(using child care in t). N = number of observations = 867.

* For women who reported working full-time in a given period after the third year, we imputed a child care value equal to 1; if the mother reported working part-time, we imputed a child care value equal to .5. Finally, if the mother does not work in a given period, we imputed a child care value of .5 if the predicted probability of child care use based on this model exceeds .65. We choose this threshold to obtain a smooth trend of child care use since childbirth and until the end of the fifth year.

† Additional controls: marital status at child birth (never married, separated, divorced, widowed), urban/rural residence, and mother's age at birth.

Appendix C

Description of the Instrumental Variables Procedure

Let S_3, S_4, S_5, and S_6 be the child's test scores at ages 3–6, respectively.[38] For example, S_3 can be the PPVT score at age 3, while S_5 could be the PIAT-Math score at age 5.[39]

Next, let Y_3, Y_4, and Y_5 represent the endogenous variables that appear in the test score equation (9) in years 3, 4, and 5 after childbirth, respectively. For example, Y_5 would include cumulative child care use up through age 5. Finally, let $R_1, R_2, \ldots, R_5$ represent vectors of instruments that are relevant for the mother's decisions in years 1–5 after the birth of the child. For example, R_5 would include welfare policy rules operative in the state of residence of the mother in year 5.[40]

Note that Y_5 is potentially influenced by $R_1 \ldots R_5$. Thus, the first-stage regressions in the 2SLS procedure will look like this:

$$
\begin{aligned}
Y_{3i} &= \alpha_0 + \alpha_1 R_{1i} + \alpha_2 R_{2i} + \alpha_3 R_{3i} + (\alpha_4 \times 0) + (\alpha_5 \times 0) + \underline{\alpha_6} X_{3i} + \varepsilon_i,\\
Y_{4i} &= \alpha_0 + \alpha_1 R_{1i} + \alpha_2 R_{2i} + \alpha_3 R_{3i} + \alpha_4 R_{4i} + (\alpha_5 \times 0) + \underline{\alpha_7} X_{4i} + \varepsilon_i, \qquad \text{(C1)}\\
Y_{5i} &= \alpha_0 + \alpha_1 R_{1i} + \alpha_2 R_{2i} + \alpha_3 R_{3i} + \alpha_4 R_{4i} + \alpha_5 R_{5i} + \underline{\alpha_8} X_{5i} + \varepsilon_i,
\end{aligned}
$$

where X_{ti} is a vector of exogenous characteristics of mothers and children that include all variables described in table 2 and $\underline{\alpha_6}$ is an associated parameter vector. Notice that $R_1, \ldots, R_t$ all enter the equation for Y_t. From (C1), we obtain the fitted values $\hat{Y}_{ti}$ for $t = 3, 4, 5$.

Finally, the second-stage regressions in the 2SLS procedure would look like this:

$$
\begin{aligned}
S_{3i} &= \beta_0 + \beta_1 \hat{Y}_{3i} + \underline{\beta_2} X_{3i} + \xi_i,\\
S_{4i} &= \beta_0 + \beta_1 \hat{Y}_{4i} + \underline{\beta_2} X_{4i} + \xi_i,\\
S_{5i} &= \beta_0 + \beta_1 \hat{Y}_{5i} + \underline{\beta_2} X_{5i} + \xi_i,\\
S_{6i} &= \beta_0 + \beta_1 \hat{Y}_{5i} + \underline{\beta_2} X_{6i} + \xi_i,
\end{aligned} \qquad \text{(C2)}
$$

where β_1 is the parameter of interest. Notice that the test score at age 6 is only influenced by the endogenous variable dated at $t = 5$ (i.e., cumulative child care use up through age 5), since at age 6 the child is of school age so day care is no longer necessary.

In the baseline specification, in order to avoid proliferation of parameters, we estimate a constrained version of the first-stage regressions (C1) where we assume the effects of the instruments on the endogenous variable

[38] Recall that cognitive ability test scores are available as early as age 3 in the NLSY.

[39] Since we have quarterly data, a test score at 3 literally means a test score in the twelfth quarter after the birth of the child.

[40] It would also include interactions of the policy rules with mother's education and AFQT.

Y_{ti} are the same in every year after birth, that is, $\alpha_1 = \alpha_2 = \ldots = \alpha_5$.[41] Finally, we constrain $\underline{\alpha}_6$, $\underline{\alpha}_7$, and $\underline{\alpha}_8$ to differ only in that a subset of the elements of X are interacted with child age.[42]

References

Anderson, T. W., Naoto Kunitomo, and Yukitoshi Matsushita. 2005. New light from old wisdoms: Alternative estimation methods of simultaneous equations in microeconomic models. CIRJE Discussion Paper no. F-321, University of Tokyo.

Anderson, T. W., and Herman Rubin. 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20:46–63.

Andrews, Donald W. K., and James H. Stock. 2006. Inference with weak instruments. In *Advances in economics and econometrics: Theory and applications; Ninth World Congress of the Econometric Society*, ed. Richard Blundell, Whitney Newey, and Torsten Persson, chap. 6. New York: Cambridge University Press.

Ben-Porath, Yoram. 1967. The production of human capital and the life cycle of earnings. *Journal of Political Economy* 75, no. 4, pt. 1:352–65.

Bernal, Raquel. 2008. The effect of maternal employment and child care on children's cognitive development. *International Economic Review* 49, no. 4:1173–1209.

Bernal, Raquel, and Michael P. Keane. 2010. Quasi-structural estimation of a model of child care choices and child cognitive ability production. *Journal of Econometrics* 156, no. 1:164–89.

Blau, David M. 1999a. The effects of child care characteristics on child development. *Journal of Human Resources* 34, no. 4:786–822.

———. 1999b. The effect of income on child development. *Review of Economics and Statistics* 81, no. 2:261–76.

Blau, David, and Janet Currie. 2004. Preschool, day care, and afterschool care: Who's minding the kids? In *Handbook of the economics of education*, ed. Erik Hanushek and Finis Welch. New Holland: Elsevier.

Blau, Francine D., and Adam J. Grossberg. 1992. Maternal labor supply and children's cognitive development. *Review of Economics and Statistics* 74, no. 3:474–81.

Cameron, Stephen V., and James J. Heckman. 1998. Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males. *Journal of Political Economy* 106, no. 2:262–333.

Carneiro, Pedro, and James J. Heckman. 2002. The evidence on credit

[41] In one robustness test we interacted all the instruments with child age, but it makes little difference.

[42] These are workbef, EXPBEF, urban, and age of mother. Other interactions were insignificant.

constraints in post-secondary schooling. *Economic Journal* 112, no. 482: 705–34.

Chase-Lansdale, P. Lindsay, Robert A. Moffitt, Brenda J. Lohman, Andrew J. Cherlin, Rebekah Levine Coley, Laura D. Pittman, Jennifer Roff, and Elizabeth Votruba-Drzal. 2003. Mothers' transitions from welfare to work and the well-being of preschoolers and adolescents. *Science* 299:1548–52.

Connolly, Sara, John Micklewright, and Stephen Nickell. 1992. The occupational success of young men who left school at sixteen. *Oxford Economics Papers* 44, no. 3:460–79.

Cragg, John G., and Stephen G. Donald. 1993. Testing identifiability and specification in instrumental variables models. *Econometric Theory* 9, no. 2:222–40.

Cunha, Flavio, and James J. Heckman. 2008. Formulating and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources* 43, no. 4:738–82.

Currie, Janet, and Duncan Thomas. 1995. Does head start make a difference? *American Economic Review* 85, no. 3:341–64.

———. 2001. Early test scores, school quality and SES: Long run effects on wage and employment outcomes. *Worker Wellbeing in a Changing Labor Market* 20:103–32.

Duncan, Greg, and NICHD Early Child Care Research Network. 2003. Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development* 74, no. 5:1454–75.

Fang, Hanming, and Michael P. Keane. 2004. Assessing the impact of welfare reform on single mothers. *Brookings Papers on Economic Activity*, no. 1:1–95.

Fuller, Wayne A. 1977. Some properties of a modification of the limited information estimator. *Econometrica* 45, no. 4:939–53.

Goodfellow, Joy, and Judy Laverty. 2003. *Grandcaring: Insights into grandparents' experiences as regular child care providers.* Canberra: Early Childhood Australia.

Gregg, Paul, Elizabeth Washbrook, Carol Propper, and Simon Burgess. 2005. The effects of a mother's return to work decision on child development in the UK. *Economic Journal* 115: F48–F80.

Hansen, Christian, Jerry Hausman, and Whitney Newey. 2008. Estimation with many instrumental variables. *Journal of Business and Economic Statistics* 26, no. 4:398–422.

Hansen, Kirstine, and Denise Hawkes, 2007. The relationship between childcare and cognitive outcomes for children of working mothers. Working paper, Centre for Longitudinal Studies, Institute of Education, University of London.

Harmon, Colm P., and Ian Walker. 1998. Selective schooling and the

returns to the quantity and quality of education. Unpublished manuscript, Department of Economics, University College Dublin.
Haveman, Robert, and Barbara Wolfe. 1995. Succeeding generations: On the effects of investments in children. *Journal of Policy Analysis and Management* 14, no 4: 639–40.
Heckman, James J., and Edward Vytlacil. 2005. Structural equations, treatment effects and econometric policy evaluation. *Econometrica* 73, no. 3:669–738.
Hutchinson, Dougal, Hilary Prosser, and Peter Wedge. 1979. The prediction of educational failure. *Educational Studies* 5, no. 1:73–82.
James-Burdumy, Susanne. 2005. The effect of maternal labor force participation on child development. *Journal of Labor Economics* 23, no. 1:177–211.
Keane, Michael P., and Kenneth I. Wolpin. 1997. The career decisions of young men. *Journal of Political Economy* 105, no. 3:473–522.
———. 2001. The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review* 42, no. 4: 1051–1103.
———. 2002. Estimating welfare effects consistent with forward looking behavior. *Journal of Human Resources* 37, no. 3:600–622.
Lamb, Michael E. 1996. Effects of nonparental child care on child development: An update. *Canadian Journal of Psychiatry* 41:330–42.
Leibowitz, Arleen. 1974. Home investments in children. *Journal of Political Economy* 82, no. 2, pt. 2: S111–S131.
Liu, H. Haiyong, Thomas Mroz, and Wilbert van der Klaauw. 2003. Maternal employment, migration, and child development. Unpublished manuscript, University of North Carolina at Chapel Hill.
Love, John M., Peter Z. Schochet, and Alicia L. Meckstroth. 1996. Are they in any real danger? What research does—and does not—tell us about child care quality and children's well-being. *Mathematica*, no. ED145 030.
McCartney, Kathleen. 1984. The effect of quality of day care environment upon children's language development. *Developmental Psychology* 20: 244–60.
Melhuish, Edward C., Ann Mooney, Eilis Hennesy, and Sue Martin. 1992. Characteristics of child care in early childhood and child development in middle childhood. Paper presented at the European Conference on Developmental Psychology, Seville, September.
Millward, Christine. 1998. Family relationships and intergenerational exchange in later life. Working Paper no. 15, Australian Institute of Family Studies, Melbourne.
Moffitt, Robert. 1992. Incentive effects of the U.S. welfare system: A review. *Journal of Economic Literature* 30, no. 1:1–61.
Murnane, Richard J., John B. Willet, and Frank Levy. 1995. The growing

importance of cognitive skills in wage determination. *Review of Economics and Statistics* 77, no. 2:251–66.

Neal, Derek A., and William Johnson. 1996. The role of premarket factors in black/white wage differences. *Journal of Political Economy* 104, no. 5:869–95.

NICHD Early Child Care Research Network. 2000. The relation of child care to cognitive and language development. *Child Development* 71, no. 4:960–80.

Pungello, Elizabeth P., Frances A. Campbell, and W. Steven Barnett. 2006. Poverty and early childhood educational intervention. Policy brief, Center on Poverty, Work, and Opportunity Policy, University of North Caroline School of Law, Chapel Hill.

Robertson, Donald, and James Symons. 2003. Do peer groups matter? Peer group versus schooling effects on academic attainment. *Economica* 70, no. 277:31–53.

Rosenzweig, Mark R., and Paul Schultz. 1983. Estimating a household production function: Heterogeneity, the demand for health inputs, and their effects on birth weight. *Journal of Political Economy* 91, no. 5: 723–46.

Ruhm, Christopher J. 2002. Parental employment and child cognitive development. *Journal of Human Resources* 39, no. 1:155–92.

Stock, James H., and Motohiro Yogo. 2004. Testing for weak instruments in linear IV regression. Working paper, Department of Economics, Harvard University.

Todd, Petra E., and Kenneth I. Wolpin. 2007. The production of cognitive achievement in children: Home, school and racial test score gaps. *Journal of Human Capital* 1, no. 1:91–36.

Zax, Jeffrey S., and Daniel I. Rees. 1998. Environment, ability, effort and earnings. Working Paper no. 9801, Center for Research on Economic and Social Policy, Denver.

 DOI: 10.1086/659343

Appendix D from Bernal/Keane, "Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers"

(JOLE, vol. 29, no. 3, p. 459)

Detailed Literature Review

This appendix gives a more detailed review of the literature than is provided in the main text.

The Effect of Maternal Employment and Child Care on Children's Cognitive Outcomes

Many prior studies, mostly in the developmental psychology literature, have used the NLSY to assess effects of maternal employment and child care use on child cognitive development. Recent reviews of this literature include Haveman and Wolf (1995), Lamb (1996), Love, Schochet, and Meckstroth (1996), Blau (1999), Ruhm (2002), and Blau and Currie (2004). Less than half of these studies provide results that are interpretable in terms of effects of specific inputs.[43] Most present simple correlations between inputs and child outcomes and do not control for family and/or child characteristics. Furthermore, some of these studies use small samples, often nonrandomly selected. In most cases, no control for selection of children into child care (or of mothers into employment) is implemented.[44]

Table D1 summarizes recent papers that use the NLSY data to assess effects of maternal employment on child cognitive outcomes. Clearly the evidence is inconclusive. Approximately a third of the studies report positive effects of maternal employment, a third report negative effects, and the rest report effects that are either insignificant or that vary by the group studied or the timing of inputs. A similar picture is seen in table D2, which summarizes recent papers that evaluate the effects of child care (and/or child care quality) on child outcomes.[45] Again, effects range from positive to negative, they are often insignificant, and they vary by group.

Reasons for this diversity of results may include the wide range of specifications that are estimated and the fact that many studies fail to control for endogeneity of employment and child care. To make our exposition of the literature clearer, it is useful to have a specific framework in mind. Consider the following equation, interpretable as a cognitive ability production function:

$$\ln S_{ijt} = \alpha_1 T_{ijt} + \alpha_2 C_{ijt} + \alpha_3 G_{ijt} + \alpha_4 X_{ijt} + \mu_j + \delta_{ij} + \varepsilon_{ijt}. \tag{D1}$$

Here S_{ijt} is a cognitive outcome (i.e., test score) for child i of mother j at age t. The log is typically taken as test scores are positive. Variable T_{ijt} is a measure of the maternal time input up through age t. This might be a scalar, as in a cumulative specification, or a specification where only current inputs matter, or a vector, if inputs at different ages are allowed to have different effects. Similarly, C_{ijt} is a measure of nonmaternal time inputs (i.e., child care) and G_{ijt} represents goods and services used in the production of child ability. Next, X_{ijt} is a set of controls for the child's initial skill endowment. This may include variables such as mother's age, education, AFQT score, and so forth (meant to capture the inherited ability endowment) and initial characteristics of the child such as gender, race, and birth weight. Turning to the error components, μ_j and δ_{ij} are family and child effects, which capture parts of the unobserved skill endowment of the child. Finally, ε_{ijt} is a transitory error term that may be interpreted as measurement error inherent in the test plus error in recording test results, along with shocks to the cognitive development path.

[43] Some studies show associations between clusters of child care arrangements/attributes and child development instead of assessing the impact of each input (Howes and Rubenstein 1985; Peterson and Peterson 1986; Studer 1992). In some cases, coefficient estimates or signs are not provided by authors (e.g., Howes and Rubenstein 1981).

[44] See, e.g., Parcel and Menaghan (1990) and Burchinal et al. (1995).

[45] Since the literature contains fewer studies of child care, table D2 is not restricted to studies that use NLSY data only.

While this general setup seems to underlie, at least implicitly, most papers in the literature, none actually estimate equation (D1), and many estimate equations that seem quite far from it. One fundamental problem is that the maternal time input T and the goods inputs G are not directly observed. Most papers ignore the problem that T is unobserved and simply use maternal employment or child care use in place of maternal time.[46]

Similarly, most papers simply ignore G, while a few proxy for it using household income or the NLSY's "HOME" environment index. The latter is problematic since it is based not just on goods inputs (e.g., books in the home) but also on maternal time inputs (e.g., time reading to the child). Baydar and Brooks-Gunn (1991) estimate effects of both maternal employment and child care but do not include goods/services. Desai, Chase-Lansdale, and Robert (1989) use maternal employment to proxy for T, average number of child care arrangements during the first 3 years after childbirth to proxy for C, and household income to proxy for G. However, as noted by Rosenzweig and Schultz (1983), Rosenzweig and Wolpin (1994), and Todd and Wolpin (2003, 2007), it can be difficult to interpret production function estimates when proxies are used for key inputs. To our knowledge, only James-Burdumy (2005) discusses the relationship between her estimating equation and a child ability production function by pointing out the difficulty of interpreting estimates when proxies are used for maternal time and goods inputs. We discuss this issue in more detail in Section IV of the main text and in appendix E.

Second, most papers in the literature have estimated specifications that include only current inputs. This is a strong assumption, especially in light of the tradition in the human capital literature of letting cumulative inputs matter. One could think of the effect of inputs cumulating over time or having a more general specification according to which the whole history of inputs since childbirth matters for the child's outcome at time t. Most papers do not discuss the implications of their assumptions regarding timing of inputs.[47] We also discuss this issue in Section IV, and we test for the importance of lagged inputs in Section VI.B and appendix G.

Finally, most papers estimate equation (D1) by ordinary least squares, ignoring potential endogeneity of inputs, that is, the potential correlation of maternal work and child care use decisions and goods inputs with the unobserved ability endowments μ_j and δ_{ij}. A few recent studies try to overcome this problem by either (i) using a very extensive set of variables to proxy for unmeasured endowments; (ii) using child or family fixed effects, or "value added" models;[48] and/or (iii) using instrumental variables.

Consider, first, the studies that can be classified as using extensive controls for the child's skill endowment. Among others, Baydar and Brooks-Gunn (1991), Vandell and Ramanan (1992), Parcel and Menaghan (1994), Han, Waldfogel, and Brooks-Gunn (2001), and Ruhm (2002) use an extensive set of observable characteristics of the child and the mother, including mother's AFQT score. In spite of this, the results of these papers are inconclusive. For example, Ruhm (2002) finds significant negative effects of maternal employment on math scores, while Parcel and Menaghan (1994) report small positive effects of maternal employment on child cognitive outcomes. Baydar and Brooks-Gunn (1991) find that maternal employment in the child's first year negatively affects cognitive outcomes, while Vandell and Ramanan (1992) find positive effects of early maternal employment on math achievement and also of current maternal employment on reading achievement.

Next, consider the studies that use fixed effects. Chase-Lansdale et al. (2003) use child fixed effects models to assess the effect of maternal employment on child outcomes. They analyzed 2,402 low-income families during the recent era of welfare reform. Their results suggest that mothers' transitions off welfare and into employment did not cause negative outcomes for preschoolers. They note, however, that this approach does not account for endogeneity of these transitions, and they do not attempt to use welfare rules as instruments for maternal employment as we do here.

James-Burdumy (2005) estimated household FE models using 498 sibling children in the NLSY. Her results suggest that effects of maternal employment vary depending on the particular cognitive ability assessment used and the timing of employment.[49] The use of sibling differences eliminates the mother (or household) fixed effect

[46] Also, most papers use one or the other of these variables and do not examine both. For example, Vandell and Ramanan (1992) estimate the effect of maternal employment on child's cognitive outcomes but do not include child care time as an additional input, while Caughy, DiPietro, and Strobino (1994) do the reverse.

[47] Notable exceptions are Blau (1999) and Duncan and NICHD (2003). Some papers use maternal employment (or child care use) at different years after childbirth but do not discuss implications of their choice in terms of properties of the underlying production function (e.g., Baydar and Brooks-Gunn 1991; Vandell and Ramanan 1992; and Waldfogel, Han, and Brooks-Gunn 2002).

[48] In the value-added approach, the test score in period t (S_{ijt}) is a function of the outcome in period $t-1$ and the inputs in period t, the idea being that the lagged test score proxies for the child's ability at the start of a period.

[49] According to James-Burdumy (2005)'s fixed effects (FE) estimates in her table 5, an increase in maternal work from 0 to 2,000 hours in year 1 of a child's life reduces the PIAT math score (measured at ages 3–5) by $(-.00117)(2{,}000) = -2.34$ points. This is similar to the effect we estimate for 1 year of full-time work (-2.1%). But she finds no significant effect of maternal employment after the first year. In contrast, we find maternal time is just as important in years 2+.

μ_j from (D1) but does not eliminate the child fixed effect δ_{ij}. It is plausible that mothers make time compensations for children depending on their ability type. Using household fixed effects does not solve this problem as maternal employment is then correlated with the sibling-specific part of the cognitive ability endowment. In addition, the FE estimator requires that input choices are unresponsive to prior sibling outcomes. If inputs to child *i* are responsive to outcomes for child *i*, then ε_{ijt} will be correlated with those inputs.

Blau (1999) and Duncan and NICHD (2003) both study the effects of child care use and quality on child outcomes. They use similar methodologies, including a wide range of proxies for unmeasured child ability (e.g., mother's AFQT and education), controls for many aspects of the home environment, and use of fixed effects and value added specifications. The main difference is that Blau (1999) uses the NLSY, while Duncan uses the NICHD Study of Early Child Care. Blau (1999) concludes that child care inputs during the first 3 years of life have little impact on child outcomes, while Duncan finds modest positive effects of improved child care quality.

From our perspective, a key difficulty in interpreting the Blau and Duncan results is that their specifications do not allow one to estimate the effect of maternal time per se. Both studies include the HOME environment index, which contains both goods inputs, such as books in the home, and also time inputs, such as how often the child is read to, eats meals with parents, or talks with the mother while she does housework. Thus, the coefficients on maternal work or child care capture the effects of those variables holding HOME fixed, which means holding some maternal time inputs fixed. In contrast, we are interested in the total impact of maternal time on child outcomes, including how a decline in the time input (from increased work or child care use) affects time reading to the child and so on.

Finally, Currie and Thomas (2001) use the NLSY to look specifically at how preschool affects outcomes. Using sibling differences and extensive controls for ability endowments, they estimate that a year of Head Start increases PPVT scores by roughly 7%, while other types of preschool have no effect. The Head Start effect persists for whites, but it is wiped out by age 10 for blacks.

The Blau, Duncan-NICHD, and Currie-Thomas papers all contain useful discussions of the limitations of fixed effects and value-added specifications. As these authors point out, neither approach is a panacea for dealing with unobserved child ability as each relies on assumptions that can be stronger than OLS. For example, the household FE estimator requires that input choices be unresponsive to the child-specific part of the ability endowment. The value-added model faces the problem that estimates of lagged dependent variable models are inconsistent in the presence of fixed effects like μ_j and δ_{ij}.[50] Neither approach, nor child fixed effects, deals with endogeneity arising because current inputs may respond to lagged test scores. An IV approach is needed to deal with this problem.

To our knowledge, only two prior papers have attempted to use IV in this context. These are Blau and Grossberg (1992) and James-Burdumy (2005).[51] Both look at effects of maternal work on child outcomes, and do not examine effects of child care use per se. More important, both papers suffer from the problem that the instruments are extremely weak. As a result, the standard errors on the maternal work variables in their two-stage least square (2SLS) regression are so large that no plausibly sized effect could possibly be significant (i.e., in each case, to attain 5% significance, maternal work over a 3 year period would have to change a child's tests scores by roughly 50 points or three standard deviations).[52] Thus, we would argue that their attempts to implement IV were not successful. Similarly, Currie and Thomas (2001) report that they attempted to use IV but could not find sufficiently powerful instruments.

The main advantage of our approach is that the welfare policy and local demand instruments that we employ are much stronger. Indeed, the first-stage marginal R^2 values we obtain using these instruments (i.e., about .09) are fairly large, and in the second stage the standard error on child care does not "explode" when these instruments are used.

Bernal (2008) takes a different approach by estimating a structural model of work and child care choices of married women. She estimates a child cognitive ability production function—which includes mother's work and

[50] Estimation of a first-differenced version of the value-added specification would eliminate the fixed effects, but Blau (1999) points out that this is difficult or impossible due to limitations of existing data. This would require three outcome observations and two lagged input observations. Even if feasible, this approach would entail a severe efficiency loss.

[51] James-Burdumy's preferred specification uses sibling differences to control for household FE and does not use IV. However, she notes that maternal employment may be endogenous in the differenced equation due to correlation of the time-varying parts of the errors in the child outcome and maternal employment equations. Of course, another source of endogeneity is correlation between unobserved child ability and the mother's decisions about work and child care.

[52] For Blau and Grossberg (1992), who use work experience prior to childbirth to instrument for maternal employment, compare cols. 1 and 2 of their table 2. For James-Burdumy (2005), who uses the percentage of the county labor force employed in services to instrument for maternal employment, compare cols. FE and IV-FE from her table 3.

child care use as inputs—jointly with the mother's work and child care decision rules, thus implementing a selection correction. Her results suggest rather sizable effects of maternal employment and child care use on child cognitive ability. In particular, 1 full year of maternal work and child care use causes a 1.8% reduction in test scores of children ages 3–7.[53]

It is interesting to extend this work to single mothers for several reasons. The first is to see if results generalize. Second, single mothers are of special policy relevance as welfare reform led to large increases in their work/child care use. Third, welfare rules have large effects on work/child care use by single mothers, so as instruments they provide a strong basis for identification. It is difficult to find plausibly exogenous variables that affect the behavior of married women so strongly.

The Relationship between Test Scores and Subsequent Outcomes (Wages, Education, and So Forth)

Several studies have examined the relationship between test scores as early as age 7 and subsequent outcomes like educational attainment and wages. This research finds that measures of cognitive achievement recorded in childhood are strong predictors of a variety of outcomes later in life. This highlights the importance of understanding what determines ability of individuals at early stages of life, particularly for the design of public policy aimed at improving labor market outcomes. We summarize some of these studies in this section.

First, consider studies using U.S. data. In the NLSY, Neal and Johnson (1996) find that scores at ages 14–21 on the Armed Forces Qualifying Test (AFQT), an IQ-type measure, are highly significant predictors of wages at ages 26–29. Murnane, Willett, and Levy (1995) use two longitudinal surveys of high school seniors to document a strong relationship between their math test scores and wages at age 24. Zax and Rees (1998) use the Wisconsin Longitudinal Study (WLS) to document that age 17 IQ is a strong predictor of wages at ages 35 and 53.

The studies linking test scores at the earliest ages to later outcomes use the British National Child Development Study (NCDS). Hutchinson, Prosser, and Wedge (1979) use the NCDS to link test scores at age 7 with scores at age 16. Similarly, Connolly, Micklewright, and Nickell (1992) find a significant positive relationship between test scores at age 7 and earnings at age 23 (in a sample of young men who left school at age 16). More recently, Robertson and Symons (2003) and Harmon and Walker (1998) find a positive association between age 7 test scores and earnings at age 33. In addition, Currie and Thomas (2001) show that a one standard deviation increase in age 16 math scores is associated with a 14% higher wage rate and a 7% higher employment rate at age 33 (for low- or medium-SES individuals). In addition, they provide evidence that age 7 (math) test scores are strong predictors of age 16 math test scores.

From our perspective, a limitation of these studies is they all look at test scores at age 7 or older (age 14 or older in the U.S. case). Do tests scores at even earlier ages predict later achievement? In appendix A we present evidence from the NLSY that PPVT scores at age 4 and PIAT reading and math scores at ages 5–6 are significantly correlated with educational attainment of youth who are at least 18 years old. For example, consider a 1 point increase in the math score at age 6 (i.e., roughly a 1% increase, as the mean score is 99.7). Holding parental background variables such as the mother's education fixed, this is associated with increased educational attainment (measured at age 18 or later) of approximately .019 years. Similarly, a 1 point (roughly 1%) increase in the reading score at age 6 is associated with an increase in highest grade completed of approximately .025 years. These estimated impacts are fairly substantial. For example, our estimates imply that a year of full-time maternal work and informal child care use reduces test scores by roughly 2.6%. This translates into an effect on completed schooling of roughly .050 to .065 years, which is a large effect.[54]

A striking aspect of the appendix A results is that mother's AFQT score is not a significant predictor of completed education. Thus, child test scores, even at ages 4–6, are better predictors of later outcomes than mother's scores. For example, in the equation that includes the child's age 6 PIAT-math score, a one standard deviation increase in the math score is predicted to raise completed education by $(.0191)(11.7) = .223$ years. In contrast, a one standard deviation increase in mother's AFQT score is predicted to increase completed education

[53] Liu, Mroz, and van der Klaauw (2003) also adopt a structural approach to estimate effects of maternal employment and school inputs on test score outcomes for 5–15-year-olds in the NLSY. They also find a negative effect of maternal employment on child outcomes. Obviously, the focus in Bernal (2008) and here is rather different since we are interested in preschool inputs.

[54] The following back-of-the-envelope calculation helps put these figures in perspective. Say people are of two types, those who finish high school (12 years) and those who finish college (16 years), and that 20% finish college. To increase average completed schooling by .06 years, the percentage finishing college must increase to 21.5%, a 7.5% increase.

by only (.00128)(18.3) = .023 years. So the point estimates imply an effect of the child's math score 10 times greater than that of the mother's AFQT score. Furthermore, in the equation that includes the child's age 6 PIAT-reading score, the coefficient on mother's AFQT is essentially zero.[55]

Table D1. The Effect of Maternal Employment on Children's Cognitive Ability

Studies Using NLSY Data	Sample	Method	Effect of Mother's Employment
Mott (1991)	2,387 1–4-year-olds	OLS	Negative
Harvey (1999)	3–12-year-olds	OLS	Negative
Ruhm (2002)	3–6-year-olds	OLS	Negative
Han, Waldfogel, and Brooks-Gunn (2001)	462 of ages birth to 8 years	OLS	Negative
Bernal (2008)	529 3–7-year-olds	Structural model	Negative
Liu, Mroz, and Van der Klaauw (2003)	5–15-year-olds	Structural model	Negative
Vandel and Ramanan (1992)	1,889 2nd graders	OLS	Positive
Parcel and Menaghan (1994)	768 3–6-year-olds	OLS	Positive
Greenstein (1995)	2,040 4–6-year-olds	OLS	Insignificant
Moore and Driscoll (1997)	1,154 5–14-year-olds	OLS	Insignificant
James-Burdumy (2005)	498 3–4-year-olds	FE and IV-FE*	Differing depending on test used
Waldfogel, Han, and Brooks-Gunn (2002)	1,872 of ages birth to 8 years	OLS and FE	Differing depending on group
Desai, Chase-Lansdale, and Robert (1989)	503 4-year-olds	OLS	Differing depending on group
Baydar and Brooks-Gunn (1991)	572 4-year-olds	OLS	Differing depending on timing
Blau and Grossberg (1992)	8,784 3–4-year-olds	OLS and IV†	Differing depending on timing
Todd and Wolpin (2007)	6–13-year-olds	FE	Not reported

*Household FE, and instruments are local market conditions (e.g., county unemployment rate and percentage of the labor force in the services sector).
†Work experience prior to childbirth is the instrument for maternal employment.

Table D2. The Effect of Child Care on Children's Cognitive Ability

Study	Sample	Method	Effect of Child Care Use
Baydar and Brooks-Gunn (1991)	572 4-year-olds	OLS	Negative (varies with timing)
Desai, Chase-Lansdale, and Robert (1989)	503 4-year-olds	OLS	Negative (only for boys)
Vandell and Corasaniti 1990	236 8-year-olds	OLS	Negative
Thornburg et al. (1990)	835 kindergarteners	OLS	Insignificant
Ackerman-Ross and Khanna (1989)	3-year-olds, whites	OLS	Insignificant
Parcel and Menaghan (1990)	697 3–6-year-olds	OLS	Insignificant
Studer (1992)	95 children	OLS	Insignificant
Burchinal et al. (1995)	6–12-year-olds	OLS	Insignificant
Blau (1999)	2,000+ 3–5-year-olds	OLS and FE*	Differing depending on quality
Caughy, DiPietro, and Strobino (1994)	867 5–6-year-olds	OLS	Differing depending on background
Dunn (1993)	4-year-olds, middle-class	OLS	Differing depending on quality of day care
Clarke-Stewart, Gruber, and Fitzgerald (1994)	2–4-year-olds, middle class	OLS	Differing depending on quality of day care
Ruopp et al. (1979)	1,600 preschool children	Experiment†	Differing depending on measure of quality
Currie and Thomas (1995)	3,477 of ages 4+	Siblings FE	Positive for Head Start
Peisner-Feinberg et al. (2001)	773 4–8-year-olds	OLS	Positive (for high-quality day care)
NICHD Early Child Care Research Network (2000)	595 0–3-year-olds	OLS	Positive for center-based arrangements
Duncan and NICHD (2003)	1,162 24–54-month-olds	OLS and FE‡	Positive effects (for high-quality day care)

*Household fixed effects.
†The National Day Care Study randomly assigned children to classrooms with different staff-child ratios and to teachers with different levels of training. However, the 64 day care centers were not randomly selected.
‡Child fixed effects.

[55] At earlier ages, the point estimates still imply larger effects of the child score than the mother's score, although the difference is not as great. For example, at age 5, a one standard deviation increase in the child's PIAT-R score is predicted to increase completed education by (.0096) (15.3) = .147 years, while a one standard deviation increase in the mother's AFQT is predicted to increase completed education by (.0045) (18.3) = .082 years. Similarly, in the equation that contains the child's PIAT-M score at age 5, the analogous figures are .091 vs. .069 years.

 DOI: 10.1086/659343

Appendix E from Bernal/Keane, "Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers"
(JOLE, vol. 29, no. 3, p. 459)

Interpreting Estimates of the Child's Cognitive Ability Production Function

As with any estimation method, the interpretation of our IV estimates of the effect of child care on child cognitive outcomes depends on a number of maintained assumptions. In this appendix we provide a more detailed discussion of those assumptions and discuss how their violation could lead to difficulties in interpreting the estimates.

In the human capital production framework (see Ben-Porath 1967) current and past inputs interact with an individual's genetic ability endowment to generate human capital. Leibowitz (1974) first used this framework to examine how investments in children add to preschool stocks of human capital. The acquisition of preschool human capital is analogous to the acquisition of human capital through schooling or on-the-job training except that at preschool ages inputs are generated by joint parental/child decisions (e.g., child tastes presumably affect parent input choices), not by choices of the child alone. Here, we focus on the cognitive ability component of human capital.

Let A_{it} be child i's cognitive ability t periods after birth. We write a production function:

$$\ln A_{it} = A(\tilde{T}_{it}, \tilde{G}_{it}, \tilde{C}_{it}, \omega_i), \tag{E1}$$

where $\tilde{T}_{it}$, $\tilde{G}_{it}$, and $\tilde{C}_{it}$ are vectors of period-by-period inputs of maternal time, goods, and child care time, respectively, up through period t, and ω_i is the child's ability endowment. Goods inputs may include nutrition, books, toys that enhance cognitive development, and the like. Child care inputs capture contributions of alternative care providers' time to child cognitive development. These may be more or less effective than mother's own time. For example, care in a group setting may contribute to child development by stimulating interaction with other children, learning activities at preschool, and so forth.

Several difficult issues arise in estimation of (E1). First, estimation of a completely general specification where inputs may have a different effect at each age t and where the endowmentω_i may differentially affect ability at each age is infeasible due to proliferation of parameters.[56] Thus, we obviously need to restrict how inputs enter (E1).

One simplification, familiar from the human capital literature, is to assume that (i) only cumulative inputs matter, rather than their timing, and (ii) the effect of the permanent unobservable is constant over time (e.g., in a Mincer earnings function, only cumulative education and experience affect human capital and the unobserved skill endowment has a constant effect). We first consider a specialization of (E1) that adopts these assumptions, and we will consider some feasible relaxations later. Letting $\breve{X}_{it} = \sum_{\tau=1,t} X_{it}$ be the cumulative amount of input X up through time t, and assuming that cumulative inputs affect lnA_{it} linearly, we obtain a special case of (E1) that takes the form

$$\ln A_{it} = \alpha_0 + \alpha_1 \breve{T}_{it} + \alpha_2 \breve{C}_{it} + \alpha_3 \ln \breve{G}_{it} + \omega_i. \tag{E2}$$

We now consider problems of estimating the production function in the special case of (E2).[57]

The second issue we face is the selection (or endogeneity) problem that arises because inputs may be

[56] For instance, if the effect of just one input is allowed to differ between every pair of input and output periods t and t', and we examine outcomes for 20 quarters after birth, we obtain $(20 \times 21)/2 = 210$ parameters for that input alone.

[57] Letting cumulative goods enter in log form is analytically convenient for reasons that will become apparent later.

correlated with the child ability endowment ω_i. To clarify this problem, assume the ability endowment is given by the equation

$$\omega_i = \beta_0 + \beta_1 Z_i + \hat{\omega}_i, \tag{E3}$$

where Z_i is a vector of mother/child characteristics correlated with the child ability endowment (e.g., mother's education and AFQT score, child gender, and birth weight) and $\hat{\omega}_i$ is the part of the ability endowment that is mean independent of observed mother and child characteristics. Next, as an illustration, assume a mother's decision rule for child care time at time t, C_{it}, can be written as

$$C_{it} = \pi_0 + \pi_1 Z_i + \pi_2 \hat{\omega}_i + \pi_3 cc + \pi_4 R_{it} + \varepsilon^c_{it},$$

where cc is the price of child care (assumed constant),[58] R_{it} is a set of welfare rules facing the mother at time t, and ε^c_{it} is a stochastic term subsuming tastes for child care use (both permanent and transitory taste shocks) and shocks to child care availability and the mother's offered wage rate. The presence of $\hat{\omega}_i$ in the decision rule means that $\breve{C}_{it}$ is endogenous in (E2), and we will require instruments that affect C_{it} still are uncorrelated with $\hat{\omega}_i$ and ε^c_{it}. Below we argue that the welfare rules R_{it} can plausibly play this role.

The third key issue in estimating (E2) is measurement of maternal time and goods inputs. One can imagine a model where mothers decide how much "quality" time to devote to the child while at home (e.g., children's time is divided between child care, "quality" time with the mother, and time spent watching TV while the mother does housework). But, we do not observe actual contact time between mothers and children, let alone how much is "quality" time. So, as is typical in the literature as a whole, we simply sidestep the issue by assuming that $T_{it} = T - C_{it}$, where T is total time in a period. Thus, we distinguish between only two types of time (i.e., time with the mother and time in child care). Then, we can rewrite (E2) as

$$\ln A_{it} = \alpha_0 + (\alpha_1 T)t + (\alpha_2 - \alpha_1)\breve{C}_{it} + \alpha_3 \ln \breve{G}_{it} + \omega_i. \tag{E4}$$

Thus, we can only estimate $\alpha_2 - \alpha_1$, the effect of time in child care relative to that of mother's time.

An issue we abstract from here is that maternal work time may influence how much of $T - C_{it}$ is "quality time." For example, a mother who uses child care but does not work might devote more of $T - C_{it}$ to "quality time." Thus, maternal work time might enter the production function directly, independently of how it affects the goods input (via the budget constraint) or how it affects C_{it}. However, for single mothers it is very difficult to address this issue because child care and maternal work time are extremely highly correlated ($\rho = .94$).[59] Thus, attempts to include both in the model fail due to severe colinearity.

The fourth key issue in estimation of (E2) is that goods inputs G_{it} are largely unobserved. For example, the NLSY contains information on books in the home but not on nutrition, health care, tutors, recreation, and the like. This problem of missing inputs plagues the entire literature, not just our study (see, e.g., Todd and Wolpin [2007] for a discussion).

To deal with this missing input problem, consider a decision rule for the cumulative goods input into the child's ability, conditional on observed mother/child characteristics Z_i (which affect permanent income and preferences), the child's ability endowment $\hat{\omega}_i$,[60] cumulative income since childbirth, child age (t), and child care usage decisions, given by

$$\ln \breve{G}_{it} = \gamma_0 + \gamma_1 Z_i + \gamma_2 \hat{\omega}_i + \gamma_3 \ln \breve{I}_{it}(W, H; R) + \gamma_4 t + \gamma_5 \breve{C}_{it} + \varepsilon^g_i, \tag{E5}$$

where the stochastic term ε^g_i captures the mother's taste for investment in the form of goods.[61] The notation $\breve{I}_{it}(W, H; R)$ highlights the dependence of income on wages, hours of market work, and the welfare rules R that determine how benefits depend on income.

Equation (E5) can be interpreted as a conditional decision rule or demand function, obtained in stage 2 of an optimization process, where, in stage 1, a mother chooses child care time C and hours of market work H, and, in

[58] That the price of child care cc is assumed constant over mothers/time is not an accident. A key problem confronting the literature on child care is that the geographic variation in cc seems too modest to use it as an IV for child care usage.

[59] Obviously, single mothers must use child care to work, and most cannot afford day care otherwise. In contrast, for married women, use of child care while not working is fairly common (see Bernal 2008).

[60] Note that the child's ability endowment may matter for two reasons: either the mother may choose good inputs based on the child's ability (e.g., she may buy educational toys to compensate a child who is having certain learning problems) or a child's ability affects the types of inputs the child demands (e.g., a high-ability child may request more books).

[61] This would arise due to heterogeneous preferences for child quality. Therefore ε^g_{it} may also be influenced by the child's tastes.

stage 2, she chooses G. (Note that the temporal aspect here is purely artificial, as in a two-stage budgeting solution.) Equation (E5) can also be interpreted as a linear approximation to the decision rule that would be generated by several alternative models of investment, both static and dynamic.[62] The key thing captured by (E5) is that a mother's decisions about goods inputs into child development may be influenced by (i.e., made jointly with) her decisions about hours of work and child care. Substituting (E5) and (E3) into (E4), we obtain

$$\begin{aligned}\ln A_{it} =\ & \alpha_0 + (\alpha_1 T)t + (\alpha_2 - \alpha_1)\breve{C}_{it} \\ & +\alpha_3[\gamma_0 + \gamma_1 Z_i + \gamma_2 \hat{\omega}_i + \gamma_3 \ln \breve{I}_{it} + \gamma_4 t + \gamma_5 \breve{C} + \varepsilon_i^g] + \beta Z_i + \hat{\omega}_i \\ =\ & (\alpha_0 + \alpha_3\gamma_0) + (\alpha_1 T + \alpha_3\gamma_4)t + (\alpha_2 - \alpha_1 + \alpha_3\gamma_5)\breve{C}_{it} \\ & +\ \alpha_3\gamma_3 \ln \breve{I}_{it} + (\beta + \alpha_3\gamma_1)Z_i + (1 + \alpha_3\gamma_2)\hat{\omega}_i + \alpha_3\varepsilon_i^g \\ =\ & \varphi_0 + \varphi_1 t + \varphi_2 \breve{C}_{it} + \varphi_3 \ln \breve{I}_{it} + \varphi_4 Z_i + \hat{\hat{\omega}}_i + \hat{\varepsilon}_i^g.\end{aligned} \tag{E6}$$

Equation (E6) is estimable because all the independent variables are observable. However, we must be careful about the appropriate estimation method and interpretation of the estimates. As we have noted, child care utilization may be correlated with the unobserved part of the child ability endowment $\hat{\omega}_i$. Furthermore, child care use may be correlated with $\hat{\varepsilon}_i^g$, the unobserved taste shifter in equation (E5) if tastes for child care use ε_{it}^c are correlated with tastes for goods investment.[63] Then, estimation of (E6) using OLS is not appropriate.

To our knowledge, it has not been previously noted that consistent estimation of an equation such as (E6) requires instruments that not only are uncorrelated with the unobserved part of the child's skill endowment, $\hat{\omega}_i$, but also are uncorrelated with the mother's tastes for goods investment in the child, ε_i^g. In order for the welfare rule parameters R_{it} to be valid instruments for cumulative child care in estimating (E6), they must be uncorrelated with these two error components. This seems like a plausible exogeneity assumption.[64] We would make a similar argument for local demand conditions.

The cumulative income variable in (E6) is also potentially endogenous, for multiple reasons. First, income depends on the jointly made child care use and work decisions. Hence, it is potentially correlated with child ability for the same reasons as were operative for child care usage. Second, income depends on the mother's wage rate, which depends on her ability endowment. To the extent that this ability endowment is not perfectly captured by Z_i, and the residual part is correlated with the child ability endowment, this will also generate correlation between the mother's income and $\hat{\omega}_i$. Thus, we need to instrument for mother's income as well. Again, we will argue that welfare rules R_{it} and local demand conditions provide plausibly valid instruments as they have important effects on work decisions yet are plausibly uncorrelated with child ability endowments.

Assuming that instrumental variables provides consistent estimates of (E6), it is important to recognize that the child care "effect" that is estimated is $\varphi_2 = \alpha_2 - \alpha_1 + \alpha_3\gamma_5$. This is the effect of child care time (α_2) relative to the effect of mother's time (α_1) plus the effect of any change in goods inputs that the mother may choose as a result of using child care ($\alpha_3\gamma_5$). In the main text we assume that $\gamma_5 = 0$, so that the estimated child care effect is in fact $\alpha_2 - \alpha_1$, the effect of child care time relative to maternal time. This is equivalent to assuming that child care time does not alter the marginal utility of investment in child quality via goods, so that C drops out of (E5). Here we want to consider the implications of failure of this assumption.

If $\gamma_5 \neq 0$, it leads to some important limitations for IV estimates of (E6). First, there is the issue that interpretation is more subtle. We must always bear in mind that we are estimating $\alpha_2 - \alpha_1 + \alpha_3\gamma_5$, which includes the behavioral response term $\alpha_3\gamma_5$, and thus we are not estimating the pure technological effect of child care.

Second, there is the issue that γ_5 is a reduced form parameter. Thus, even given a consistent estimate of $\varphi_2 = \alpha_2 - \alpha_1 + \alpha_3\gamma_5$, we must be careful to use it only to evaluate the impact of changes in child care induced by

[62] For example, if $\gamma_3 = 0$, there is a fixed rate of investment determined by permanent characteristics and the cumulative goods input grows at a rate given by γ_4. At the other extreme, if $\gamma_1 = \gamma_2 = \gamma_4 = 0$ and $\gamma_3 = 1$, then demand for goods is simply proportional to current income ($G_{it} = \exp(\gamma_0)I_{it}$).

[63] For instance, a mother with a high taste for child quality may both spend more time with the child (i.e., use less day care) and invest more in the child in the form of goods. This would tend to bias estimated effects of child care usage in a negative direction since not only the maternal time input but also the goods input is lower for children in child care.

[64] In app. J, we report means of child test scores prior to 1990 by state, broken down by whether the state subsequently implemented welfare waivers (i.e., moved toward welfare reform early) and by whether the state implemented strict or lenient welfare rules after 1996. There is no significant difference in average prereform test scores between "strict" and "lenient" states.

policy changes that leave γ_5 unchanged. Some policies will do this, while others will not. Of course, situations where parameters may be invariant to some policies and not others are common in structural estimation—see, for example, Keane and Wolpin (2001) for a discussion of this issue in the context of their model of college attendance decisions, which can only be used to simulate policies that leave parents' decision rule for providing financial support to children unaffected.

Here, if $\gamma_5 \neq 0$, then, in interpreting estimated effects of child care on child outcomes, we must be careful to view them as applying only to policy experiments that do not alter the decision rule for goods investment in children (E5). As this decision rule is conditional on work and child care usage decisions, it will be invariant to policies that leave the budget constraint conditional on those decisions unchanged. A work requirement that induces a woman to work and use child care but that leaves her wage rate and the cost of care unaffected would fall into this category.

In contrast, a policy such as child care subsidies would shift the budget constraint conditional on work and child care usage, so it is unlikely to leave (E5) unchanged. Such subsidies would not only alter child care use, but potentially also goods inputs, and in a way not captured by $\alpha_3\gamma_5$.

To our knowledge, the issues we are discussing here were first raised by Rosenzweig and Schultz (1983). They call an equation like (E6), where proxy variables are substituted for one or more unobserved inputs, a "hybrid" production function, and they discuss the potential problems that may arise in interpreting estimates of such a function. Rosenzweig and Wolpin (1994) and Todd and Wolpin (2003, 2007) also discuss this issue. It is important to note, however, that there is no ideal way out of this problem. The alternative to proxying for unobserved goods inputs is to simply ignore them, which would lead to omitted variable bias. As noted by Todd and Wolpin (2007), it is not obvious a priori which approach would lead to greater bias.

 DOI: 10.1086/659343

Appendix F from Bernal/Keane, "Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers"

(JOLE, vol. 29, no. 3, p. 459)

Tests of the Pooling Hypothesis

It is desirable to pool the PPVT, PIAT-M, and PIAT-R tests for use in estimation of equation (9) in the text since this leads to a substantial efficiency gain (i.e., due to increased sample size). However, before pooling the data it is important to test that the production function parameters are invariant across the three tests. We consider two types of test of the pooling hypothesis.

The first type of test is based on differences in scores on two different tests for the same child at the same age. Let S_{it}^1 and S_{it}^2 denote scores of child i on two different tests, both measured at the same age t (e.g., test 1 could be the PPVT and test 2 could be the PIAT-M, both recorded at age 5). Let W_{it} denote the complete set of variables included in equation (9). Then, if we run the regression

$$\ln S_{it}^1 - \ln S_{it}^2 = W_{it}\Gamma + \varepsilon_{it} \tag{F1}$$

under the null hypothesis that all parameters of (9) are equal for each test, we should not be able to reject the hypothesis that $\Gamma = 0$. More generally, let $W_{it} = (U_{it}, X_{it})$, where U_{it} is a subset of variables whose effects are invariant across tests and X_{it} is a subset of variables whose effects differ by test. Then, if we run the regression

$$\ln S_{it}^1 - \ln S_{it}^2 = U_{it}\Delta + X_{it}\Upsilon + \varepsilon_{it}, \tag{F2}$$

we should not be able to reject the null hypothesis that $\Delta = 0$.

A second type of test involves pooling the data on all three tests together and estimating a fully interacted specification of the form:

$$\ln S_{it}^j = W_{it}\Omega + W_{it}d_{it}\Gamma + \varepsilon_{it}, \tag{F3}$$

where $W_{it}d_{it}$ is a complete set of interaction terms between W_{it} and a set of test indicators d, with associated coefficient vector Γ. Under the pooling hypothesis we should not be able to reject the null hypothesis that $\Gamma = 0$. More generally, suppose that only the subset of variables X_{it} have effects that differ by test. Then, if we run the regression

$$\ln S_{it}^j = U_{it}\Psi + U_{it}d_{it}\Delta + X_{it}\Phi + X_{it}d_{it}\Upsilon + \varepsilon_{it}, \tag{F4}$$

we should not be able to reject the null hypothesis that $\Delta = 0$. That is, once we allow the subset of variables X_{it} to have effects that differ by test, we should not be able to reject the hypothesis that the remaining variables U_{it} have common effects.

Results of these tests are reported in table F1. The first four rows report results for tests of the first type, that is, regressions of test score differences on all regressors in equation (9). Note that we can only examine four pairs of tests because not all tests are observed at all ages (i.e., only the PPVT and not the PIATs are taken at ages 3 and 4). The first column contains tests of the type F1, where no conditioning variables are allowed.

For the PIAT math and reading tests at ages 5 and 6, we cannot reject the null hypothesis of common coefficients at the 5% level (p-values of .308 and .055, respectively). However, for the difference between the PPVT at age 5 and either the PIAT math or reading at age 5, we do reject the hypothesis of common coefficients (p-values of .025 and .018, respectively).

Thus, we turn to the second column of table F1, which contains tests of the form F2. Here, we include four conditioning variables in X_{it}: race (i.e., nonwhite) interacted with two test dummies (PIAT-R is the omitted category) as well both AFQT and the AFQT missing indicator interacted with the PPVT dummy. These variables were chosen because they were significant in the PPVT-PIAT regressions. As we see in column 2, with these variables included in X_{it}, we cannot reject the null hypothesis that all other coefficients are zero (i.e., $\Delta = 0$) at the 10% level.

The last row of table F1 reports tests of the second type—regressions where we fully interact all variables in equation (9) with test indicators, allowing their effects to differ freely by test. The first column reports tests of the type F3, where we test if all the interaction terms are jointly significant. The hypothesis that all interactions are zero (i.e., $\Gamma = 0$) is clearly rejected ($p = .0000$).

However, in the second column we report tests of the type F4, where we include in X_{it} the same four controls as before: interactions of race (i.e., nonwhite) with test dummies and interactions of AFQT and AFQT missing with the PPVT dummy. With these controls included in X_{it}, we can no longer reject the null hypothesis that all other test interactions are zero (i.e., $\Delta = 0$) at the 5% level.

Thus, we conclude that in our baseline model it is appropriate to pool the three tests provided interactions of race (i.e., nonwhite) with test dummies and interactions of AFQT and AFQT missing with the PPVT dummy are included in the model.

Table F2 contains a comparison of our baseline results with and without these interactions. Note that failure to include them leads to some upward bias in the cumulative child care coefficient, from −0.53% per quarter to −0.73% per quarter.[65]

The interaction coefficients are interesting in themselves. The coefficient on the AFQT with PPVT interaction is .0022, with a *t*-statistic of about 5.5. As the standard deviation of AFQT is 18.30, this implies that a one standard deviation increase in the mother's AFQT raises the PPVT score relative to the PIAT scores by (.0022)(18.30) = 4%. The race coefficients imply that, ceteris paribus, whites and nonwhites have essentially identical scores on the PIAT-R but that nonwhites have scores 11.5% lower on the PPVT and 1.9% lower on the PIAT-M. Why PPVT scores are more sensitive to mother's race and AFQT than PIAT scores is a very interesting question for future research, but it goes beyond the scope of the present investigation.

Table F3 reports results where we let the cumulative child care coefficient differ by test. (Otherwise this model is exactly the same as our baseline specification in table F2 that includes all the interaction terms noted above.) We report results for LIML using the 14 factors as instruments.[66] The point estimates imply that the cumulative child care effects are greater for the PIAT-M and PIAT-R (−0.50% and −0.61% per quarter, respectively) than they are for the PPVT (−0.32% per quarter). And indeed, while the coefficients for the PIAT tests are significant at the 5% level, that for the PPVT is not. However, the χ^2 test for the hypothesis that the three child care coefficients are equal is 1.53, which has a *p*-value of .464. Thus, there is no clear evidence of differential effects of child care by test. Based on these results, we conclude that it is appropriate to pool the tests (provided the interaction terms noted above are included in the model).

Finally, we repeated the above analysis breaking the nonwhite indicator into separate black and Hispanic indicators. This produced almost identical results. That is, given interactions of both black and Hispanic with the test dummies (and the AFQT/PPVT interaction), we could no longer reject pooling. Having separating indicators for blacks and Hispanics increased the estimated child care effect only slightly from −0.53% to −0.57% per quarter.[67] Also, table 12 reports results from a model that lets child care effects differ by race. We do not find significant differences.

[65] Note that the model without interactions uses 16 factors as instruments, while the model with interactions uses only 14 factors. In each case the factors were chosen using the method described in Sec. VI.C and the candidate factors were identical. But to retain a factor as an instrument we required that it have a *t*-statistic of at least 3 in at least one of the first-stage regressions for the three endogenous variables. Once the interaction terms were added to the first-stage regressions, two of the factors no longer satisfied the $t > 3$ cutoff, so the number of IVs dropped from 16 to 14. However, results obtained using all 16 instruments in both models were essentially identical.

[66] Results using the full set of 78 excluded instruments are quite similar.

[67] It is not surprising that separating out blacks and Hispanics has little effect on the results because the estimated test dummies for blacks and Hispanics are fairly similar. For blacks, PPVT scores are 13.8% lower than for whites, ceteris paribus. For Hispanics, the figure is 10.2%. For the PIAT-M, the analogous figures are 2.8% and 2.1%. For the PIAT-R, blacks and whites have essentially identical scores, while the scores for Hispanics are 2.7% lower.

Table F1. Tests of the Pooling Hypothesis: Significance of Test Differences

Dependent Variable	*p*-Value for *F*-Test of Joint Significance of All Variables Except: None	AFQT and RACE
Test difference (at same age):		
PPVT5-MATH5	.0251	.1082[a]
PPVT5-READ5	.0181	.1793[a]
MATH5-READ5	.3083	...[b]
MATH6-READ6	.0553	...[b]
All tests pooled[c]	.0000	.0537

[a]AFQT × dPPVT, afqtmiss × d PPVT, RACE × dPPVT, and RACE × dMATH.
[b]The entire set of control variables are insignificant, so controls for AFQT and RACE not necessary.
[c]Presents *p*-value for *F*-test of joint significance of all explanatory variables interacted with test dummies (except for what is stated in the column heading).

Table F2. Tests of the Pooling Hypothesis

	Baseline Excludes Interactions of AFQT, Nonwhite with Test Dummies[a]	Baseline Includes AFQT, Nonwhite Interacted with Test Dummies[b]
Cumulative child care	−.00727*	−.00533*
	(.0026)	(.0025)
Log(cumulative income)	.02809	.01062
	(.0246)	(.0266)
No. of children	−.03036*	−.02545*
	(.0068)	(.0064)
I[mother's age<20]	.02225[+]	.02368*
	(.0119)	(.0116)
I[mother's age>30]	.00071	.00603
	(.0260)	(.0256)
Mother's education	.01282*	.01297*
	(.0031)	(.0030)
Mother's AFQT score	.00129*	.00066*
	(.0003)	(.0003)
AFQT score × dPPVT		.00220*
		(.0004)
I[AFQT missing]	.05796*	.02919
	(.0194)	(.0181)
I[AFQT missing] × dPPVT		.09809*
		(.0358)
I[nonwhite]	−.03999*	.00451
	(.0111)	(.0113)
I[nonwhite] × dPPVT		−.11986*
		(.0208)
I[nonwhite] × dMATH		−.02376*
		(.0102)
Male	−.02566*	−.02593*
	(.0071)	(.0069)
Birthweight	.00448	.00555
	(.0063)	(.0063)
I[work before]	.02998*	.02674*
	(.0125)	(.0125)
I[work before] × skill	.00894	.00988
	(.0118)	(.0118)

Experience before childbirth	.00616	.00565
	(.0066)	(.0067)
Experience × mother's age	−.00014	−.00013
	(.0002)	(.0002)
I[never married]	.01807	.01696
	(.0332)	(.0318)
I[separated]	.03902	.03414
	(.0340)	(.0326)
I[divorced]	.04024	.03218
	(.0355)	(.0341)
I[urban]	.02188+	.02450*
	(.0116)	(.0114)
Child's age	.03554*	.03817*
	(.0113)	(.0122)
dPPVT	−.25327*	−.19561*
	(.0103)	(.0249)
dMATH	−.07792*	−.05796*
	(.0040)	(.0092)
R^2	.3507	.3844
k[c]	1.009	1.005
Weak/many-instruments test	16.45	15.33

NOTE.—Dependent variable is log (test score). N = number of observations = 3,787. Estimation method = LIML. Robust standard errors (Huber-White) by child clusters are in parentheses.

[a]Instruments are 16 factors derived from the factor analysis of our original 78 instruments described in the note to table 5 in the main text.

[b]Instruments are 14 factors derived from the factor analysis of our original 78 instruments.

[c]k is the parameter of the k-class estimator, which equals one for 2SLS and exceeds one for LIML.

+Significant at the 10% level.

*Significant at the 5% level.

Table F3. Child Care Effects by Test

Estimation Method	Baseline Includes AFQT, NONWHITE Interacted with Test Dummies
LIML with 14 factors:	
Child care × dPPVT	−.00318
	(.00361)
Child care × dMATH	−.00495*
	(.00241)
Child care × dREAD	−.00611*
	(.00258)
χ^2 test for identical effects by test	1.53
	(.464)

NOTE.—Dependent variable is log (test score). Standard errors are in parentheses.

*Significant at the 5% level.

 DOI: 10.1086/659343

Appendix G from Bernal/Keane, "Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers"

(JOLE, vol. 29, no. 3, p. 459)

Testing the Cumulative versus Current Child Care Specifications

Table G1 compares estimates of specifications where current test scores depend on cumulative or current child care. The cumulative child care variable sums up quarters of child care (with a quarter of full-time care counting as 1 and a quarter of part-time care counting as 0.5). In the cumulative specification in column 1, the estimated cumulative child care coefficient is $-.00533$ (with SE $= .0025$). This means that each additional quarter of full-time child care reduces test scores by roughly 0.53%. This corresponds to an effect of -2.1% per year.

In the current specification in column 2, the estimated effect of current child care is -0.0269 (SE $= .0137$). This implies that a single quarter of full- time child care reduces test scores by 2.7%. This very large effect estimate is not surprising given the very strong persistence in child care use. The probability a mother uses child care in quarter t conditional on having used it in quarter $t - 1$ is 93.5%. Similarly, the own transition rate for nonuse of child care is 89.1%.[68] Thus, current child care use provides a very good proxy for lagged child care use.

We can formally test the cumulative specification versus the current specification in the following way. First, we can reject the current specification if lagged child care inputs matter conditional on current inputs. Second, the cumulative specification implies the restriction that the current and lagged child care coefficients are all equal. If we find that lagged inputs matter but that coefficients are not equal, it implies that a more general dynamic model is appropriate.

If we add 4 years of lagged child care indicators to the current specification, the p-value for their joint significance is .0329. Thus, we reject that only current child care matters. Furthermore, a χ^2 test for equality of coefficients on current and lagged child care gives a p-value of .1748. This supports the cumulative specification. Admittedly, however, a cumulative specification would be hard to reject given the great persistence in child care use noted above. As a result of that persistence, lagged child care indicators are highly collinear and individual lag coefficients are very imprecisely estimated. This is why we do not report coefficients on the individual lags.

[68] These transition rates are calculated only over the 12 quarters after birth for which we observe child care use directly.

Table G1. Cumulative versus Current Specification

	LIML		OLS	
	Cumulative Specification (1)	Current Specification (2)	Cumulative Specification (3)	Current Specification (4)
Cumulative child care	−.00533*		.00098	
	(.0025)		(.0008)	
Current child care		−.02690*		.00395
		(.0137)		(.0027)
Log(cumulative income)	.01062	.01503	−.00324	−.00311
	(.0266)	(.0275)	(.0057)	(.0056)
Mother's education	.01297*	.01346*	.01051*	.01047*
	(.0030)	(.0032)	(.0026)	(.0026)
Mother's AFQT	.00066*	.00059⁺	.00059*	.00060*
	(.0003)	(.0003)	(.0002)	(.0002)
R^2	.3844	.3682	.3994	.3995
Root MSE	.1729	.1752	.1714	.1714
k[a]	1.005	1.005		
Weak/many-instruments test	15.33	13.29		

NOTE.—Dependent variable is log(test score). N = number of observations = 3,787. Instruments are 14 factors derived from the factor analysis of our original 78 instruments described in the note to table 5 in the main text. Robust standard errors (Huber-White) by child clusters are in parentheses.
[a] k is the parameter of the k-class estimator, which equals one for 2SLS and exceeds one for LIML.
*Significant at the 5% level.
⁺Significant at the 10% level.

Appendix H from Bernal/Keane, "Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers"
(JOLE, vol. 29, no. 3, p. 459)

Results Using Principal Components as Instruments

Our estimation procedure involves factor analyzing the instruments, rotating the factors by the commonly used varimax method, and then estimating each factor using the commonly used regression method. We then run regressions of the endogenous variables on the factors, and choose the subset of factors that have the greatest explanatory power for the endogenous variables. This led us to choose 14 factors that had *t*-statistics of at least 3 in those regressions.

The factor rotation serves two valuable purposes in this procedure: First, it gives us orthogonal factors, thus leading to low correlations among our instruments. Second, it gives us factors with a clear interpretation (e.g., factor 6 measures welfare benefit levels) so we can judge if the factors have sensible signs in the first stage regressions.

An alternative procedure is to simply use principal components factor analysis and choose factors that have eigenvalues above some cutoff level (e.g., a commonly used eigenvalue threshold for retaining a factor is 1). We could then estimate these factors, and use the estimated principal factors as instruments. (Note: Factors must always be estimated in a second stage after factor analysis is performed, regardless of whether one uses rotated factors or principal components. The estimated factors are simply weighted sums of the original instruments.)

In our view there are several fundamental problems with this principal factor approach. First, note that the most important principal factors (i.e., those with the largest eigenvalues) are the ones that explain most of the covariance among the instruments. This is not at all the same thing as the set of factors that will best explain the endogenous variables.

To see the difference, consider the following simple example. Suppose we factor analyze nine variables. The first eight are highly correlated among each other, while the ninth has low correlations with the first eight. A factor analysis of these data produces two factors. The first accounts for the covariances among the first eight variables. It has a very large eigenvalue because it explains most of the covariance in the data. There will also be a second factor, but it has a very small eigenvalue because only the ninth variable loads on it (i.e., it explains little of the covariance in the data). Now, this scenario is perfectly consistent with a situation where the ninth variable, and hence the second factor, has far more explanatory power for the endogenous variables than does the first factor. But a researcher choosing instruments using a principal factors/largest eigenvalues criterion might mistakenly conclude that only the first factor is important and discard the relatively unimportant second factor—leading to a very inefficient IV procedure.

As the above example illustrates, factors will have large eigenvalues simply because they capture correlations among a large number of variables regardless of whether those variables have much explanatory power for the endogenous variables. Our own results give a very clear illustration of this phenomenon. Note that it is the factor with the sixth largest eigenvalue that has the most explanatory power for child care usage. This factor captures welfare benefit levels.

In contrast, the factor with the largest eigenvalue is the one that captures time limits. Why? Simply because we use eight variables to measure time limits and only two variables to measure benefit levels. So of course the factor that captures time limits has a much larger eigenvalue because it captures a much larger set of covariances among the instruments. But, as Fang and Keane (2004) note, time limits are not nearly as important as benefit levels in explaining welfare participation.

Despite these arguments, as an experiment, we decided to compare our results to those obtained using unrotated principal factors as instruments. The results are reported in table H1. First, we report results using an eigenvalue cutoff of 1.0 to retain factors. This leads us to retain 13 factors, which explain 91.8% of the covariance among the original 78 instruments. Estimates of these factors are then used as instruments. Given these instruments, the estimate of the child care effect is −0.38% per quarter, and it is not significant.

Reducing the eigenvalue cutoff to 0.5 leads us to retain 18 factors, which explain 96% of the covariance among the original 78 instruments. This gives an estimated child care effect of −0.51% per quarter, which is significant at the 10% level.

Finally, reducing the eigenvalue cutoff to 0.3 leads us to retain 21 factors, which explain 97.3% of the covariance among the 78 instruments. This gives an estimate of the child care effect of −0.63% per quarter, which is significant at the 5% level. It is only at this point that the estimates “settle down” in the sense that adding more factors did not noticeably alter results.

The last row of table H1 reports our baseline results using 14 selected rotated factors. Three advantages of our procedure are apparent. Our 14 factors produce a more efficient estimate (i.e., a SE of .0025 vs. .0027 when we use the 21 principal factors). Furthermore, our 14 factors have a higher correlation with the endogenous variables than do the 21 principal factors (e.g., a Shea partial R^2 for child care of .0967 vs. .0960). We also obtain a higher value of the Cragg-Donald weak instrument statistic (15.33 vs. 11.13). Note that the third advantage actually follows directly from the first two: the weak instrument statistic is higher because we obtain higher correlations with the endogenous variables using fewer instruments.

Table H1. Sensitivity to Factors Included as Instruments

Factors	Effect of Child Care on Log Test Scores (Baseline Specification)	Eigenvalue Cutoff	Variance Explained	Shea’s Partial R^2	Cragg-Donnald Statistic
Number of principal factors:					
13	−.00375	1	.918	.0831	15.02
	(.0028)				
18	$-.00513^{+}$	.5	.960	.0925	12.27
	(.0026)				
21	−.00633*	.3	.973	.096	11.13
	(.0027)				
14 selected rotated factors	−.00533*			.0967	15.33
	(.0025)				

$^{+}$Significant at the 10% level.
*Significant at the 5% level.

 DOI: 10.1086/659343

Appendix I from Bernal/Keane, "Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers"

(JOLE, vol. 29, no. 3, p. 459)

Alternative Tests for Effect of Mother's Age at Birth of Child

As we indicated in the main text, our welfare policy instruments are correlated with mother's age at childbirth due to the timing of waivers/TANF and the structure of the NLSY79. Specifically, a child had to be born in 1990 or later to have any chance of being affected by the reform before the age of 6. Moreover, women who had children prior to 1990 tend to be younger at childbirth than those who had children later. Indeed, from 1990 onward, all births are to mothers in their twenties and thirties, while prior to 1990 many were to teenage mothers. The youngest women in the NLSY, those who were age 14 in January 1979, would be age 24 by January 1990. Thus, the large majority of mothers who would have been affected by welfare reform would have been at least age 24 at the time of childbirth.

So we have that welfare reform (*a*) positively affected maternal work/child care use and (*b*) is positively correlated with age at childbirth. This correlation alone will not generate bias. However, if (i) mother's age at birth has a positive effect on child cognitive ability and (ii) we fail to adequately control for mother's age in the main equation, this will generate a spurious positive effect of maternal work/child care use on child cognitive test scores. As we actually find a negative effect, the plausible concern is that we understate this negative effect.

Table 9, column 2, in the main text presents an alternative specification that adds additional controls for mother's age at childbirth to the main equation. Specifically, we add age and age squared (whereas the baseline model contains only dummies for whether the mother was under age 20 or over age 33). The inclusion of the additional controls has a negligible impact on the estimates (compare table 9 cols. 1 and 2). In other results (not reported) we also tried eliminating all age controls from the main equation, and we found that this also has a negligible effect (the estimated child care effect ranges from −0.49% per quarter with no controls for age at all to −0.52% with all four controls).

Strikingly, inspection of the coefficients on the mother's age controls reveals that mother's age at childbirth has little effect on child cognitive outcomes (conditional on measures of the mother's human capital like education, AFQT). Thus, the potential source of bias noted in (i) above, that is, a positive association between maternal age at birth and child cognitive outcomes, is simply not present. Indeed, the column 1 estimates imply that, ceteris paribus, children of teenage mothers have 2.4% higher test scores. Thus, if anything, it seems that (controlling for economic resources) maternal youth is slightly beneficial for child outcomes. But this effect is too weak for the presence or absence of age controls to have much impact on the child care coefficient.

To further address the age issue, in table I1 we report estimates for subsamples of women based on age at childbirth. We restrict the sample to women who were 24+, 24–34, or 24–30 years old at childbirth. For these subsamples, the estimated effects of child care range from −2.3% to −1.6% per year. However, due to the reduced sample size, these estimates are not significant. So in columns 4–6 we use the full sample, but we interact the child care coefficient with dummies for whether the mother's age at childbirth was outside the indicated range. None of the interactions is significant, and the main effect of child care is quite stable across the three columns (i.e., it ranges from −2.3% to −2.4% per year and is always significant). Thus, we find no evidence that the age/welfare policy correlation leads to significant bias in our estimates of the child care effect.

Table I1. Alternative Tests for Effect of Mother's Age at Childbirth

	(Mother's) Age Restricted Samples			Models with Age Interactions		
	24+ (1)	24–34 (2)	24–30 (3)	24+ (4)	24–34 (5)	24–30 (6)
Cumulative child care	−.00585	−.00515	−.00404	−.00575*	−.00565*	−.00608*
	(.0050)	(.0045)	(.0058)	(.0029)	(.0028)	(.0029)
Cumulative child care × mother's age[a]				.00153	.00100	.00126
				(.0018)	(.0011)	(.0406)
χ^2 test for additional age interactions (*p*-value)				.1282	.1040	.1327
No. of observations	1,680	1,643	1,345	3,787	3,787	3,787
R^2	.3758	.3780	.3915	.3872	.3857	.3854
k[b]	1.082	1.076	1.098	1.040	1.040	1.039
Weak/many-instruments test	4.25	4.49	3.85	5.68	5.72	5.72

NOTE.—Dependent variable is log(test score). Instruments are all 78 policy variables, local demand conditions, and interactions described in note of table 5. Estimation method is LIML. Robust standard errors (Huber-White) by child clusters are in parentheses.

[a]Mother's age dummy defined as the complement of the age stated in the corresponding column heading.

[b]k is the parameter of the k-class estimator.

*Significant at the 5% level.

 DOI: 10.1086/659343

Appendix J from Bernal/Keane, "Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers"

(JOLE, vol. 29, no. 3, p. 459)

Alternative Tests for State Effects

As we noted in Section VI.D, a possible concern with our results is cross-state correlation between the instruments and unobserved child skill endowments. That is, it is possible that states where children had relatively low unobserved skill endowments may have adopted stricter welfare reform. This would bias our estimated child care effect negatively.

Here we test for the existence of such a correlation. Specifically, in table J1 we group states into those that adopted more versus less strict approaches to welfare reform along five different dimensions. Then we compare average test scores in the prereform period between each group of states. In each case, there is no significant difference in average child test scores in the prereform period between states that subsequently adopted more versus less strict welfare reform programs. Thus, there is no evidence of cross-state correlation between the instruments and child skill endowments.

Table J1. Average Test Scores for Children Born Prior to 1990

	Average	SD	*t*-Test
States that implemented time-limit waivers	93.34	(1.82)	−.46
States that did not implement time-limit waivers	92.42	(1.08)	
States that implemented work-requirement waivers	89.77	(1.35)	1.56
States that did not implement work-requirement waivers	93.45	(1.09)	
States with time limits lower than 3 years	90.2	(2.46)	.87
States with time limits higher than 3 years	93.02	(1.00)	
States with immediate work requirements	93.48	(1.81)	−.66
States with work requirements of at least 1 month	92.20	(.95)	
States with age of youngest child exemption < 6 months	93.40	(2.20)	−.51
States with age of youngest child exemption > 6 months	92.38	(.84)	

SOURCE.—NLSY79, sample of single mothers.

 DOI: 10.1086/659343

Appendix K from Bernal/Keane, "Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers"

(JOLE, vol. 29, no. 3, p. 459)

Alternative Approaches to Clustering

In our main results we have clustered the standard errors by child. In our data, there are an average of 2.59 test score observations per child, and we would expect strong correlations among these scores, arising from child latent ability endowments. Indeed, if we do an analysis of variance of the residuals from our main equation, we find that child effects account for 57% of the variance of the residuals. In order to assess the importance of clustering, table K1 presents results with and without clustering for four different specifications of the instrument set.

Specifically, table K1 reports estimates for our baseline specification of the main equation estimated using (*a*) the full set of 78 instruments, (*b*) only state-level instruments, but including their interactions with mother's education and AFQT, (*c*) only state-level instruments with no interactions, and (*d*) the 14 instruments obtained via factor analysis.[69] Standard errors obtained without clustering are reported directly under the estimates, and standard errors clustered by child are reported below those. As we see, for the child care coefficient, clustering by child increases standard errors by 15%–40%, depending on the instrument set. This large increase is not surprising given the high within-cluster correlation of the errors. It is interesting that the largest increase is in the model with all 78 instruments (40%), while the smallest is in the model that uses only the 14 factors as instruments (15%).

One might argue that we should cluster at a higher level. For instance, one could argue that we should cluster at the level of the mother since 368 of our 944 mothers have multiple children and there may be unobserved mother effects that are common in the family. Or one could argue that we should cluster at the level of the state since much (but certainly not all) of the variation in our instruments is at the state level. These results are also reported in table K1.

Note that clustering by state is not possible for the models in columns 1 and 2, where the number of instruments is 78 and 63, respectively. This is because the construction of the robust covariance matrix estimator requires that there be more clusters than instruments. In columns 3 and 4, the number of instruments is 25 and 14, respectively, so it is possible to cluster by state. Even so, we would view the state-clustered standard errors with some caution since the asymptotics of the robust covariance matrix estimator relies on the number of clusters growing large relative to the number of moment conditions (instruments).

The results show that clustering by mother leads to slightly larger standard errors than clustering by child, while clustering by state leads to slightly smaller standard errors. However, these differences are minor compared to those between clustering by child versus not clustering at all. Thus, the choice of clustering level makes essentially no difference to any of our results.

Still, we prefer the child-clustered results for a number of reasons. First, the number of state clusters is too small relative to the number of instruments to have great faith in the asymptotic theory. (Indeed, we can only implement state clusters in a small subset of our specifications.) Furthermore, in an analysis of variance of the residuals, state effects account for only 2% of the variance. Thus, they seem to be quite unimportant compared to child effects, which account for 57%.

Second, regarding clustering by mother, we note that mother effects can explain 17% of the variance of the residuals. Thus, mother effects are not as important as child effects, and the latter completely subsume the former. An issue with mother effects is that the within-cluster correlation is much higher for clusters that contain

[69] The specifications correspond to table 10; col. 1, table 10, col. 7; table 10, col. 8; and table 6, col. 6, respectively.

only one child than for clusters that contain two or more children. Hence, the intracluster correlation will be a weighted average of the two, and this mongrel parameter will be too low for the former and too high for the latter. Intuitively, the larger the intracluster correlation, the greater is the downward adjustment of the sample size needed to obtain the effective sample size used to compute standard errors. Our intuition is that this will cause the single-child observations to be counted too much and the multichild observations counted too little.

Table K1. Comparison of Factor-Baseline Specification by Type of Clustering

	All 78 Instruments (1)	Only State-Specific Instruments[a] (2)	Only State-Specific IVs without Interactions (3)	14 Factors as Instruments (4)
Cumulative child care	−.00522	−.00623	−.00642	−.00533
No cluster	(.0020)*	(.0020)*	(.0020)*	(.0022)*
Child clusters	(.0028)[+]	(.0025)*	(.0026)*	(.0025)*
Mother clusters	(.0030)[+]	(.0027)*	(.0028)*	(.0027)*
State clusters			(.0024)*	(.0024)*
Log(cumulative income)	.01037	−.00300	.00529	.01062
No cluster	(.0160)	(.0163)	(.0219)	(.0192)
Child clusters	(.0242)	(.0246)	(.0361)	(.0266)
Mother clusters	(.0255)	(.0256)	(.0379)	(.0284)
State clusters			(.0358)	(.0314)
Mother's education	.01276	.01316	.01167	.01298
No cluster	(.0023)*	(.0025)*	(.0027)*	(.0025)
Child clusters	(.0032)*	(.0034)*	(.0037)*	(.0033)*
Mother clusters	(.0034)*	(.0038)*	(.0040)*	(.0033)*
State clusters			(.0041)*	(.0034)*
Mother's AFQT	.00066	.00078	.00073	.00077
No cluster	(.0003)*	(.0003)*	(.0003)*	(.0003)*
Child clusters	(.0003)*	(.0003)*	(.0004)[+]	(.0003)*
Mother clusters	(.0003)*	(.0003)*	(.0004)[+]	(.0003)*
State clusters			(.0003)*	(.0003)*

NOTE.—Dependent variable is log(test score).
[a]Excludes all individual-specific welfare rules, such as, whether a woman could have a hit a time limit or a work requirement (i.e., all variables with an *i* subscript in table 1). Standard errors are in parentheses.
[+]Significant at the 10% level.
*Significant at the 5% level.

 DOI: 10.1086/659343

Appendix L from Bernal/Keane, "Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers"

(JOLE, vol. 29, no. 3, p. 459)

Robustness Checks for Informal versus Formal Child Care Results

As we noted in Section VI.G, our finding that informal child care has adverse affects relative to formal care is arguably the most important of this study in that it may provide a rationale for government programs (e.g., CCDF in the United States or Child Care Benefit in Australia) that create incentives for mothers to use formal rather than informal care. Thus, we subjected this result to the same battery of robustness tests we applied to our estimates of the effect of child care in general. Before we begin, recall that we found (see table L1) that a year of informal care reduces test scores by 0.64% per quarter or 2.6% per year ($t = -2.22$) while formal care has a positive but insignificant effect (+0.30% per quarter or +1.2% per year; $t = 0.46$).

In table L1, we report the same battery of sensitivity tests to changes in specification of the main equation that we reported in table 9. First, consider robustness to controls for mother's age at childbirth. As we see in column 2, adding additional age controls only shifts the informal child care coefficient slightly, from −0.64% to −0.72% per quarter, and it remains significant at the 5% level. Again, the age coefficients show little effect of mother's age at childbirth on child outcomes.

In table L2, we report results analogous to table I1, where we restrict the sample to mothers who were 24+, 24–34, or 24–30 years old at childbirth. Here the loss of sample size causes the estimates to lose significance, but the point estimates in the −0.47% to −0.72% range are still broadly consistent with our earlier results.[70]

We now return to table L1 and consider the sensitivity of our results to five other changes in the specification of the main equation. First, in column 3, we drop the mother's AFQT score. This increases the estimated informal child care coefficient from −2.6% to −3.5% per year. As in table 9, it also produces a large increase in the cumulative income coefficient, which becomes highly significant. As we noted in Section VI.D, this suggests that, with AFQT omitted, transitory income becomes more important, as it proxies for the mother's permanent income/skill endowment.

In the column 4, we consider the sensitivity of our results to controls for the ages of siblings. Including separate regressors for number of children aged 0–5 and 6–17 (both treated as endogenous) has essentially no impact on the estimated informal and formal child care effects.

Next, in column 5, we consider aggregate time effects. As in table 9, when we include a quadratic time trend it has a U-shape, implying that an aggregate factor not included in our model first drove down test scores and this was followed by a recovery. But including the time trend only slightly reduces the estimated informal child care effect, from −2.6% to −2.4% per year, and the formal child care effect is little changed. Thus, any bias from omitted time effects appears to be minor. Results were essentially identical using unrestricted time dummies.

In the column 6 of table L1 we add state fixed effects to the main equation. This increases the estimated cumulative informal child care effect from −.64% per quarter to a very large value of −1.01%. But this is imprecisely estimated, and hence it is only significant at the 10% level. This large increase in the informal child care coefficient is similar to what happened in table 9 for all child care.

However, as we pointed out in Section VI.D, we are skeptical of the fixed effects estimates for several reasons. First, as we already noted, state effects account for only 2% of the variance of the residuals. Second, as we see in table L1, column 6, state effects are not jointly significant at the 10% level, so there is no clear evidence for including them. Third, even if unobserved child ability did differ by state, it would only induce bias if it were correlated with the instruments, that is, if states with low-ability children adopted stricter reforms. This

[70] In table I1 we deal with this small sample problem by also using the full sample and interacting the child care variable with indicators for mother's age at childbirth. The same approach was not successful here because now we have two interactions (age at childbirth with formal and informal care) and thus two new endogenous variables. We were not able to find reasonably powerful instruments for the interaction between formal care and mother's age at childbirth.

strikes us as implausible a priori, and appendix J presents empirical evidence that is contrary to this idea. Fourth, in the child care context, we are skeptical of the strict exogeneity assumption required for consistency of the fixed effects estimator.[71] Finally, Keane and Wolpin (2002) show that fixed effects can lead to very misleading results if expected future values of policy variables matter for current decisions.[72]

However, our fixed effects results are at least comforting in that they suggest that failure to include fixed effects does not bias our estimates of the informal child care effects in a negative direction.[73] That is, the child care coefficient becomes even more negative when state effects are included. So there is no indication that failure to account for state fixed effects is driving our result that informal care has a negative effect.

Finally, in the last column of table L1 we consider a model where test scores enter in levels. The estimates imply that a quarter of informal child care reduces scores by 0.49 points. As the mean score is 91.9, this is -0.53% per quarter, or -2.1% per year. This is a bit smaller then the -2.6% effect estimated in logs. The estimate for formal care is $+0.40\%$ per quarter, but it is still insignificant.

In table L3 we consider sensitivity of our formal/informal child care results to the specification of the instrument list. This table is analogous to table 10 in the main text. It reports LIML results using the baseline list of 78 instruments in column 1 and seven variants on that list in columns 2–8. As in the main text, we must use the full set of 78 instruments rather than the 14 factors for this exercise because we want to consider the impact of dropping certain types of variables from the instrument list (and there is not a one-to-one mapping between factors and types of instruments because some factors pick up multiple aspects of welfare reform).

In column 2 we exclude CCDF spending. Excluding this instrument increases the estimated informal child care effect slightly from -2.6% to -2.8%. In column 3 we use only the main features of TANF as instruments: time limits, work requirements and disregards. This increases the estimated informal child care effect substantially to -3.8% per year.

In column 4 we drop all the TANF-related instruments, using other aspects of policy and demand environment to identify the child care effect. In column 5 we go even further and drop all instruments specific to the welfare reforms of the 1990s (e.g., TANF, CCDF, EITC), using only instruments that would have varied across states/time regardless. These are state welfare grant levels and local demand conditions (i.e., state unemployment rates and 20th percentile wage rates).

Clearly the attempts to apply IV in columns 4 and 5 were not successful. These smaller instrument sets do not contain variables that are effective at predicting whether people use formal versus informal child care. Recall that in Section VI.G we noted that mothers are more likely to use formal relative to informal care if (i) a state does not have a work requirement, (ii) it has young child or other work requirement exemptions, (iii) it has a longer work requirement time limit, (iv) work requirements were implemented more recently, (v) less time has elapsed since a time limit could have hit, (vi) remaining eligibility is greater, (vii) a state has higher CCDF spending, or (viii) earnings disregards are greater. Thus, all of the effective predictor variables are aspects of the TANF reforms or waivers. When we drop all of these variables, the point estimates for cumulative formal and informal child care become very imprecise. Furthermore, in columns 4 and 5, the Cragg-Donald weak-instrument test statistic falls to very low values of 0.81 and 0.44, respectively. Thus, we view the results in columns 4 and 5 as uninterpretable. (Note that table L4 reports on the explanatory power of the instruments in the first-stage regressions.)

Recall that in our reduced form regression, we interact all policy and demand variables with mother's education and AFQT. This allows changes in welfare policy and demand conditions to have different effects on different types of mothers (e.g., welfare rules are less important for the college educated). In column 6 we drop these interactions to see how important they are. The estimated informal child care effect is reduced only slightly to -2.5% per year versus -2.6% under the baseline. The formal child care estimate is hardly changed.

Next, recall that some of our instruments are tailored to individuals based on ages of their children (e.g., whether a woman could have reached the time limit; see Sec. III.A). In column 7 we drop these individual specific instruments and rely purely on cross-state and over-time variation to identify the child care effect. The

[71] The strict exogeneity assumption will fail if children's test score realizations at age t affect future inputs into child production and/or how the welfare policy rules evolve. (See also the discussion of this point in Sec. II.)

[72] A state fixed effect controls for a state's average level of welfare generosity. Thus, using state effects, we estimate the impact of deviations from the average level of child care use induced by deviations from average welfare rules. Such short-run effects may differ from effects of long-run policy changes, and the latter are presumably of greater interest.

[73] A negative bias would emerge if states where children had low-skill endowments adopted stricter welfare reform (making it look like stricter welfare rules led to more work/child care, which in turn led to lower scores). Recall that we looked at this issue directly in app. J. There we group states into those that adopted more versus less strict approaches to welfare reform along five different dimensions. Then we compare average test scores in the prereform period between each group of states. In each case, there is no significant difference in average child test scores in the prereform period between states that subsequently adopted more versus less strict welfare reform programs. Thus, there is no evidence of cross-state correlation between the instruments and child skill endowments.

resulting estimate of the informal child care effect is −3.1% per year, which is a bit larger than our baseline estimate. In column 8 we go further and also drop the interactions of the remaining instruments with mother's education and AFQT. This gives an estimate of −2.9% per year.

In summary, our result of a negative effect of informal child care is robust to a wide range of alternative instrument sets, with estimates ranging from −2.5% to −3.8% per year and all but one estimate between −2.5% and −3.1% (compared to our baseline of −2.6%). Our finding of no adverse effect of formal child care is also robust across all these instrument sets.

The only exception to this pattern was when we dropped instruments related to TANF and other aspects of the welfare reforms of the 1990s. These instruments are crucial for predicting whether mothers use formal or informal care. This is because, loosely speaking, reforms that had a "carrot"-like aspect (i.e., that created positive incentives to work) tended to induce more use of formal care, while reforms that took more of a "stick" approach (e.g., work requirements) tended to induce more use of informal care. Thus, without these instruments, we are unable to identify separate informal and formal child care effects.

Table L1. Robustness with Respect to the Specification of the Main Equation

	Baseline (1)	Additional Mother's Age Controls (2)	Removing AFQT (3)	Children by Ages (4)	Year Effects (5)	State Effects[a] (6)	Test Score in Levels[b] (7)
Cumulative formal child care	.00302	.00241	.00290	.00299	.00266	−.00157	.36834
	(.0066)	(.0064)	(.0085)	(.0068)	(.0068)	(.0064)	(.5131)
Cumulative informal child care	−.00643*	−.00721*	−.00867*	−.00644*	−.00610*	−.01013^{+}	−.48528*
	(.0029)	(.0029)	(.0035)	(.0029)	(.0030)	(.0061)	(.2312)
Log(cumulative income)	.00719	.00717	.05873*	.00732	−.00054	.01755	.25614
	(.0233)	(.0249)	(.0261)	(.0237)	(.0259)	(.0288)	(1.8241)
Mother's education	.01203*	.01381*	.01495*	.01200*	.01326*	.01359*	.92806*
	(.0032)	(.0034)	(.0038)	(.0033)	(.0032)	(.0036)	(.2549)
Mother's AFQT score	.00061*	.00060*		.00061*	.00062*	.00066*	.07486*
	(.0003)	(.0003)		(.0003)	(.0003)	(.0003)	(.0259)
Child's age	.04067*	.04173*	.02843*	.04068*	.04326*	.04583*	3.58776*
	(.0123)	(.0126)	(.0137)	(.0123)	(.0127)	(.0139)	(.9379)
Mother's age		−.01208	−.00696				
		(.0156)	(.0166)				
(Mother's age)^2		.00022	.00007				
		(.0003)	(.0003)				
I[age of mother$_i$<20]	.02016^{+}	.00535	.00607	.01986	.00924	.01999^{+}	1.45023
	(.0114)	(.0153)	(.0162)	(.0142)	(.0117)	(.0113)	(.9125)
I[age of mother$_i$>=33]	.01236	.00263	.01098	.01227	.00801	.01536	.88193
	(.0254)	(.0322)	(.0353)	(.0256)	(.0258)	(.0259)	(2.1840)
Mother's age at first birth		−.00163					
		(.0018)					
No. of children	−.02471*	−.02517*	−.02115*		−.02026*	−.02937*	−2.00045*
	(.0061)	(.0058)	(.0073)		(.0064)	(.0079)	(.4889)
No. children 0–5				−.02458*			
				(.0066)			
No. children 6–17				−.02497*			
				(.0103)			
Year (at time of test)					−.01052*		
					(.0038)		
(Year)^2					.00056*		
					(.0003)		
R^2	.3776	.3743	.3318	.3776	.3828	.3649	.3916
k[c]	1.039	1.037	1.047	1.038	1.041	1.040	1.036
Weak/many-instruments test	4.50	4.61	4.55	4.42	4.50	4.15	4.50

NOTE.—Dependent variable is log(test score). N = number of observations = 3,787. Estimation method is LIML. Instruments are all 78 policy variables, local demand conditions, and interactions described in note of table 5. Robust standard errors (Huber-White) by child clusters are in parentheses.

[a] The mean score is 91.9, so the point estimate implies an informal child care effect of −.53 per quarter, or −2.1% per year.

[b] Joint significance test for state FE is 28.25 (.104).

[c] k is the parameter of the k-class estimator, which equals one for 2SLS and exceeds one for LIML.

$^{+}$Significant at the 10% level.

*Significant at the 5% level.

Table L2. Robustness with Respect to Mother's Age at Childbirth

	(Mother's) Age Restricted Samples		
	24+	24–34	24–30
Cumulative formal child care	.00064	.00071	.00307
	(.0067)	(.0062)	(.0086)
Cumulative informal child care	−.00720	−.00600	−.00466
	(.0050)	(.0046)	(.0059)
Mother's education	.01102*	.01060*	.01023
	(.0048)	(.0048)	(.0063)
Mother's AFQT	.00052	.00045	.00018
	(.0005)	(.0005)	(.0006)
No. of observations	1,680	1,643	1,345
R^2	.3714	.3764	.3859
k[a]	1.080	1.074	1.096
Weak/many-instruments test	4.03	4.16	3.62

NOTE.—Dependent variable is log(test score). Instruments are all 78 policy variables, local demand conditions, and interactions described in note to table 5. Robust standard errors (Huber-White) by child clusters are in parentheses.
[a] k is the parameter of the k-class estimator.
*Significant at the 5% level.

Table L3. Robustness with Respect to the Instrument List

	Original Set of IVs[a] (1)	Excluding CCDF[b] (2)	Only TL, WR, and ED (3)	Excludes TL, WR, and ED (4)	Only BEN and Local Demand (5)	Original Set without Interactions (6)	Only State-Specific Instruments[c] (7)	Only State-Specific IVs[c] without Interactions (8)
Cumulative formal child care	.00302	.00268	.00067	−.00208	−.12307	.00335	.00367	.00572[+]
	(.0066)	(.0066)	(.0067)	(.0761)	(.5455)	(.0063)	(.0060)	(.0126)
Cumulative informal child care	−.00643*	−.00713*	−.00946*	−.00439	.01311	−.00617*	−.00765*	−.00719*
	(.0029)	(.0029)	(.0041)	(.0097)	(.0792)	(.0028)	(.0026)	(.0033)
Log(cumulative income)	.00719	.00697	.02390	.01842	.11612	−.00439	−.01152	−.01400
	(.0233)	(.0239)	(.0355)	(.0767)	(.5549)	(.0230)	(.0251)	(.0425)
Mother's education	.01203*	.01232*	.01252*	.01035	.01783	.01180*	.01270*	.01096*
	(.0032)	(.0032)	(.0034)	(.0053)[+]	(.0303)	(.0031)	(.0034)	(.0037)
Mother's AFQT	.00061*	.00062*	.00056*	.00056	.00100	.00068*	.00075*	.00074*
	(.0003)	(.0003)	(.0003)	(.0004)	(.0016)	(.0003)	(.0003)	(.0004)
R^2	.3776	.3735	.3576	.3861	...	.3766	.3634	.3568
k[d]	1.039	1.038	1.034	1.005	1.002	1.007	1.024	1.005
Weak/many-instruments test	4.50	4.59	5.13	.81	.44	5.91	3.90	2.67
p-value, overidentification test	.648	.615	.392	.968	...	.894	.786	.803
No. of instruments	78	75	58	27	18	26	63	25

NOTE.—Dependent variable is log(test score). N = number of observations = 3,787. See descriptions of instruments in table 1. Estimation method is LIML. TL = time limits; WR = work requirements; ED = earnings disregards; BEN = benefit amounts. Robust standard errors (Huber-White) by child clusters are in parentheses.
[a] All 78 policy variables, local demand conditions, and interactions described in note to table 5 in the main text. Unless otherwise noted in the column heading, all specifications include these interactions.
[b] CCDF = Child Care Development Fund expenditures.
[c] Excludes all individual-specific welfare rules, such as, whether a woman could have a hit a time limit or a work requirement (i.e., all variables with an i subscript are in table 1).
[d] k is the parameter of the k-class estimator, which equals one for 2SLS and exceeds one for LIML.
[+] Significant at the 10% level.
*Significant at the 5% level.

Table L4. Explanatory Power of Instruments in First-Stage Regressions for Informal Child Care (Instruments in Table L3)

Instruments Listed in Note to Table L3	(Partial Correlation)^2	(Shea Partial Correlation)^2	Incremental R^2	F-Statistic	p-Value
Original set of IVs	.1337	.1446	.0921	16.310	.000
Excluding CCDF	.1338	.1391	.0885	17.270	.000
Only TL, WR, and ED	.1049	.0914	.0694	21.730	.000
Excludes TL, WR, and ED	.0753	.0751	.0499	7.610	.000
Only BEN and local demand	.066	.0387	.0437	9.830	.000
Original set without interactions	.0997	.1037	.0714	9.990	.000
Only state-specific IVs	.1129	.1340	.0747	7.580	.000
State-specific IVs without interactions	.0769	.0899	.0562	5.760	.000

NOTE.—Dependent variable is cumulative informal child care. R^2 of first-stage regression with only exogenous variables = .3885.

 DOI: 10.1086/659343

Appendix M from Bernal/Keane, "Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers"

(JOLE, vol. 29, no. 3, p. 459)

Who Uses Formal versus Informal Care and Relatives versus Nonrelatives?

In this appendix we present evidence on what type of mothers are more likely to use different types of child care. In column 1 of table M1 we present a logit for whether a mother uses formal or informal child care (conditional on child care use). The results show that more educated, urban women with fewer children are more likely to use formal care. This suggests that formal care is higher quality since it is typically used by women who can afford more expensive care.

In column 2 of table M1 we present a logit for use of relatives versus nonrelatives (conditional on using informal care). It is the more educated, urban women with fewer children who are more likely to use nonrelatives. Again, this is suggestive that nonrelatives provide higher quality care than relatives.

These findings about who uses each type of care appear consistent with our finding that only informal care has adverse effects on child cognitive development, and that, among types of informal care, only care by relatives has a significant negative effect.

Table M1. Who Is Using Formal Child Care and Care Provided by Nonrelatives?

	Dependent Variable	
	Formal Child Care Used (1 If Formal; 0 If Informal) (1)	Care Provided by Nonrelative (1 If Nonrelative; 0 If Relative) (2)
Mother's education	.12126*	.11247*
	(.0149)	(.0167)
Mother's age at birth	−.01140^{+}	.02024*
	(.0056)	(.0061)
No. of children	−.08925*	−.04318*
	(.0191)	(.0213)
Urban/rural	.17590*	.63539*
	(.0637)	(.0798)
No. of observations	12,167	9,471
Pseudo R^2	.0116	.0209

Note.—Method of estimation is logit.
$^{+}$Significant at the 10% level.
*Significant at the 5% level.

 DOI: 10.1086/659343

Appendix N from Bernal/Keane, "Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers"

(JOLE, vol. 29, no. 3, p. 459)

Works Cited in Online Appendices D–M

References

Akerman-Ross, Susan, and Prabha Khanna. 1989. The relationship between high quality day care to middle-class 3-year-olds' language performance. *Early Childhood Research* 4, no. 1:97–116.

Baydar, Nazil, and Jeanne Brooks-Gunn. 1991. Effects of maternal employment and child-care arrangements on preschoolers' cognitive and behavioral outcomes: Evidence from children of the National Longitudinal Survey of Youth. *Developmental Psychology* 27, no. 6:932–45.

Ben-Porath, Yoram. 1967. The production of human capital and the life cycle of earnings. *Journal of Political Economy* 75, no. 4, pt. 1:352–65.

Bernal, Raquel. 2008. The effect of maternal employment and child care on children's cognitive development. *International Economic Review* 49, no. 4:1173–1209.

Blau, David M. 1999. The effects of child care characteristics on child development. *Journal of Human Resources* 34, no. 4:786–822.

Blau, David, and Janet Currie. 2004. Preschool, day care, and afterschool care: Who's minding the kids? In *Handbook of the economics of education*, ed. Erik Hanushek and Finis Welch, 1163–1278. Amsterdam: Elsevier.

Blau, Francine D., and Adam J. Grossberg. 1992. Maternal labor supply and children's cognitive development. *Review of Economics and Statistics* 74, no. 3:474–81.

Burchinal, Margaret R., Sharon L. Ramey, Mary K. Reid, and James Jaccard. 1995. Early child care experiences and their association with family and child characteristics during middle childhood. *Early Childhood Research Quarterly* 10, no. 1:33–61.

Caughy, Margaret, Janet A. DiPietro, and Donna M. Strobino. 1994. Daycare participation as a protective factor in the cognitive development of low-income children. *Child Development* 65:457–71.

Chase-Lansdale, P. Lindsay, Robert A. Moffitt, Brenda J. Lohman, Andrew J. Cherlin, Rebekah Levine Coley, Laura D. Pittman, Jennifer Roff, and Elizabeth Votruba-Drzal. 2003. Mothers' transitions from welfare to work and the well-being of preschoolers and adolescents. *Science* 299:1548–52.

Clarke-Stewart, Alison, Christian P. Gruber, and Linda M. Fitzgerald. 1994. Children at home and in day care. Hillsdale, NJ: Erlbaum.

Connolly, Sara, John Micklewright, and Stephen Nickell. 1992. The occupational success of young men who left school at sixteen. *Oxford Economics Papers* 44, no. 3:460–79.

Currie, Janet, and Duncan Thomas. 1995. Does Head Start make a difference? *American Economic Review* 85, no. 3:341–64.

———. 2001. Early test scores, school quality, and SES: Long-run effects on wage and employment outcomes. *Worker Wellbeing in a Changing Labor Market* 20:103–32.

Desai, Sonalde, Lindsay Chase-Lansdale, and Michael T. Robert. 1989. Mother or market? Effects of maternal employment on the intellectual ability of 4-year-old children. *Demography* 26, no. 4:545–61.

Duncan, Greg, and NICHD Early Child Care Research Network. 2003. Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development* 74, no. 5:1454–75.

Dunn, Loraine. 1993. Proximal and distal features of daycare quality and children's development. *Early Childhood Research Quarterly* 8, no. 2:167–92.

Fang, Hanming, and Michael P. Keane. 2004. Assessing the impact of welfare reform on single mothers. *Brookings Papers on Economic Activity*, no. 1:1–95.

Greenstein, Theodore N. 1995. Are the "most advantaged" children truly disadvantaged by early maternal employment? *Journal of Family Issues* 16, no. 2:149–69.

Han, Wen-Jui, Jane Waldfogel, and Jeanne Brooks-Gunn. 2001. The effects of early maternal employment on later cognitive and behavioral outcome. *Journal of Marriage and the Family* 63, no. 2:336–54.

Harmon, Colm P., and Ian Walker. 1998. Selective schooling and the returns to the quantity and quality of education. Unpublished manuscript, Department of Economics, University College Dublin.

Harvey, Elizabeth. 1999. Short-term and long-term effects of early parental employment on children of the National Longitudinal Survey of Youth. *Developmental Psychology* 35, no. 2:445–59.

Haveman, Robert, and Barbara Wolfe. 1995. Succeeding generations: On the effects of investments in children. *Journal of Policy Analysis and Management* 14, no. 4:639–40.

Howes, Carollee, and Judith L. Rubenstein. 1981. Toddler peer behavior in two types of day care. *Infant Behavior and Development* 4 (March): 387–93.

———. 1985. Determinants of toddlers' experiences in day care: Age of entry and quality of setting. *Child Care Quarterly* 14, no. 2:40–51.

Hutchinson, Dougal, Hilary Prosser, and Peter Wedge. 1979. The prediction of educational failure. *Educational Studies* 5, no. 1:73–82.

James-Burdumy, Susanne. 2005. The effect of maternal labor force participation on child development. *Journal of Labor Economics* 23, no. 1:177–211.

Keane, Michael P., and Kenneth I. Wolpin. 2001. The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review* 42, no. 4:1051–1103.

———. 2002. Estimating welfare effects consistent with forward-looking behavior. *Journal of Human Resources* 37, no. 3:600–622.

Lamb, Michael E. 1996. Effects of nonparental child care on child development: An update. *Canadian Journal of Psychiatry* 41, no. 6:330–42.

Leibowitz, Arleen. 1974. Home investments in children. *Journal of Political Economy* 82, no. 2, pt. 2:S111–S131.

Liu, H. Haiyong, Thomas Mroz, and Wilbert van der Klaauw. 2003. Maternal employment, migration, and child development. Unpublished manuscript, University of North Carolina at Chapel Hill.

Love, John M., Peter Z. Schochet, and Alicia L. Meckstroth. 1996. Are they in any real danger? What research does—and does not—tell us about child care quality and children's well-being. *Mathematica*, ED145 030.

Moore, Kristin A., and Anne Driscoll. 1997. Low-wage maternal employment and outcomes for children: A study. *Future of Children* 7, no. 1:122–27.

Mott, Frank L. 1991. Developmental effects of infant care: The mediating role of gender and health. *Journal of Social Issues* 47, no. 2:139–58.

Murnane, Richard J., John B. Willet, and Frank Levy. 1995. The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics* 77, no. 2:251–66.

Neal, Derek A., and William Johnson. 1996. The role of premarket factors in black/white wage differences. *Journal of Political Economy* 104, no. 5:869–95.

NICHD Early Child Care Research Network. 2000. The relation of child care to cognitive and language development. *Child Development* 71, no. 4:960–80.

Parcel, Toby L., and Elizabeth G. Menaghan. 1990. Maternal working conditions and children's verbal facility: Studying the intergenerational transmission of inequality from mothers to young children. *Social Psychology Quarterly* 53, no. 2:132–47.

———. 1994. Early parental work, family social capital, and early childhood outcomes. *American Journal of Sociology* 99, no. 4:972–1009.

Peterson, Carole, and Richard Peterson. 1986. Parent-child interaction and day care: Does quality of day care matter? *Journal of Applied Developmental Psychology* 7, no. 1:1–15.

Peisner-Feinberg, Ellen S., Margaret R. Burchinal, Richard M. Clifford, Mary L. Culkin, Carollee Howes, Sharon L. Kagan, and Noreen Yazejian. 2001. The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development* 72, no. 5:1534–53.

Robertson, Donald, and James Symons. 2003. Do peer groups matter? Peer group versus schooling effects on academic attainment. *Economica* 70, no. 277:31–53.

Rosenzweig, Mark R., and Paul Schultz. 1983. Estimating a household production function: Heterogeneity, the demand for health inputs, and their effects on birth weight. *Journal of Political Economy* 91, no. 5:723–46.

Rosenzweig, Mark R., and Kenneth Wolpin. 1994. Are there increasing returns to the inter-generational production of human capital? *Journal of Human Resources* 29, no. 2:670–93.

Ruhm, Christopher J. 2002. Parental employment and child cognitive development. *Journal of Human Resources* 39, no. 1:155–92.

Ruopp, Richard, Jefrrey Travers, Frederic Glantz, and Craig Coelen. 1979. *Children at the center: Final results of the National Day Care Study*. Cambridge, MA: Abt.

Studer, Marlena. 1992. Quality of center care and preschool cognitive outcomes: Differences by family income. *Sociological Studies of Child Development* 5:49–72.

Thornburg, Kathy, Peggy Pearl, Dwayne Crompton, and Jean M. Ispa. 1990. Development of kindergarten children based on child care arrangements. *Early Childhood Research Quarterly* 5, no. 1:27–42.

Todd, Petra E., and Kenneth I. Wolpin, 2003. On the specification and estimation of the production function for cognitive achievement. *Economic Journal* 113, no. 485:3–33.

———. 2007. The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital* 1, no. 1:91–36.

Vandell, Deborah L., and Mary A. Corasaniti. 1990. Variations in early child care: Do they predict subsequent social, emotional and cognitive differences? *Early Childhood Research Quarterly* 5, no. 4:555–72.

Vandell, Deborah L., and Janaki Ramanan. 1992. Effects of early and recent maternal employment on children from low-income families. *Child Development* 63, no. 4:938–49.

Waldfogel, Jane, Wen-Jui Han, and Jeanne Brooks-Gunn. 2002. The effects of early maternal employment on child cognitive development. *Demography* 39 no. 2:369–92.

Zax, Jeffrey S., and Daniel I. Rees. 1998. Environment, ability, effort, and earnings. Working Paper no. 9801, Center for Research on Economic and Social Policy, Denver.