# Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers[1]

Forthcoming *Journal of Labor Economics*, 2011

Raquel Bernal[2]
Department of Economics
Universidad de los Andes
Colombia

Michael P. Keane
ARC Federation Fellow
University of Technology Sydney
Research Fellow, Arizona State

First Version: September 15th, 2004
This Version: November 15th, 2010

**Abstract**

We evaluate the effect of childcare vs. maternal time inputs on child cognitive development using the single mothers from the NLSY79. To deal with non-random selection of children into childcare, 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/childcare use. Thus, 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 childcare reduces child test scores by 2.1%.

## 1. Introduction

The effect of parental time inputs and childcare 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 about age 16.[3] 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 childcare. In this paper, we take a small step towards learning more about development of

---

[1] Web appendices available: http://economia.uniandes.edu.co/es/profesores/planta/bernal_raquel/documentos_de_trabajo
[2] We acknowledge support from Australian Research Council grants FF0561843 and DP0774247, and Bernal acknowledges financial support from "The Searle Fund." Corresponding author: rbernal@uniandes.edu.co.
[3] See, e.g., Keane and Wolpin (1997, 2001), Cameron and Heckman (1998), Cunha and Heckman (2008).

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, 5 and 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 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 previous literature may be failure of many studies to adequately control for biases arising from two factors: (1) Women who work/use childcare may be systematically different from those who do not, both in the constraints they face and their tastes; (2) 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 also more likely to work/use childcare. Failure to control for this correlation would cause estimated effects of maternal employment (childcare) 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 childcare). Again, the estimated effect of maternal employment (childcare) 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 paper, we estimate child cognitive ability production functions for single mothers in the NLSY. We focus on single mothers because major changes in welfare rules in the 1990s led to dramatic and plausibly exogenous variation in their work incentives. 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, as well as giving States much more leeway to set local rules. Major rule changes involved benefit and work requirement time limits, earnings disregards and childcare assistance. These policy changes greatly increased employment and childcare use among single mothers with children aged 0-5. Indeed, the percent 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 use these as instruments in estimation of the cognitive ability production function. One key source of identification comes from comparing outcomes for children born before 1990 vs. those born later, as waivers and PRWORA only affected the latter group. However, benefit levels and local demand conditions – which also have good explanatory power for behavior of single mothers – differ substantially over States and time for the whole sample period.

An important technical problem arises as the welfare rules are very complex. Thus, we need many variables to characterize them. As a result, we face a "many instrument problem." That is, 2SLS estimates can be severely biased (towards OLS) when the number of over-identifying instruments is large - see, e.g., Stock and Yogo (2004), Anderson, Kunitomo and Matsushita (2005), Hansen, Hausman and Newey (2007), Andrews and Stock (2006). We deal with this problem in two ways: First, we use the LIML estimator, which corrects the 2SLS bias in the many instrument case (see Hansen, Hausman and Newey, 2007). Our 2SLS estimates fall in between LIML and OLS, suggesting 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 instrument problem for 2SLS, but 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 one additional year of childcare 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, as it is well known that in general IV 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 non-relatives) 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 childcare obtains because 75% of single mothers use informal care.

Finally, it is interesting to examine how the welfare policy changes of the mid-90s affected 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 test scores were modestly reduced.

The paper is organized as follows. In section 2 we review the relevant literature. In Section 3 we describe the welfare policy and local demand variables that we use as instruments. Section 4 derives the cognitive ability production function that we estimate. In Section 5 we describe the data and Section 6 presents the estimation results. Section 7 concludes.

## 2. Literature Review

### 2.1. 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 childcare use on child cognitive development. Recent reviews of this literature include Love, Schochet and Meckstroth (1996), Blau (1999a), Lamb (1996), Haveman and Wolfe (1995), 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 Web Appendix A. The most obvious feature of the existing literature is that it has produced very mixed results. Regarding effects of maternal employment on child outcomes, about a third of the studies report positive effects, a third negative effects, and the rest report ambiguous effects. Results on effects of childcare 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 as test scores are positive. $T_{ijt}$ 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., childcare), 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 – e.g., the mother's education, AFQT score, etc., or the child's birthweight. The error components, $\mu_j$ and $\delta_{ij}$ are family and child effects, which capture parts of the child's *unobserved* ability endowment. Finally, $\varepsilon_{ijt}$ 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 (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, simply using maternal employment or childcare 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 as 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 – i.e., correlation of maternal work and childcare use decisions, as well as goods inputs, with the unobserved ability endowments $\mu_j$ and $\delta_{ij}$. A few recent studies try to overcome this problem by: (1) using an extensive set of controls for skill endowments, (2) using child or family fixed effects, or "value added" models,[4] and/or (3) using instrumental variables.

As we note in Web Appendix A, even studies that use extensive controls for the child's skill endowment and/or 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 childcare 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 differences (as in James-Burdumy (2005)) eliminates the mother (or household) fixed effect $\mu_j$ but does not eliminate the child effect $\delta_{ij}$. Thus, the household FE estimator requires that input choices be unresponsive to the child specific ability endowment. But it is plausible that mothers make time compensations for children depending on their ability type. Also, the household FE estimator assumes 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 does not provide a panacea for dealing with unobserved child ability either. The value added model faces the problem that estimates of lagged dependent variable models are inconsistent in the presence of fixed effects like $\mu_j$ and $\delta_{ij}$.[5] And 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 the fixed effects and value added approaches do not address.

[4] 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.

[5] Estimation of a first-differenced version of the value-added specification would eliminate the fixed effects, but Blau (1999a) points out 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.

To our knowledge, only two prior papers used IV: Blau and Grossberg (1992) and James-Burdumy (2005). Both look at effects of maternal work on child outcomes, and do not examine effects of childcare *per se*. More importantly, 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 three years would have to change a child's tests scores by roughly 50 points or 3 standard deviations). Thus, their attempts to use IV were not successful.[6] The main advantage of our approach is that the instruments we employ are much stronger, as we'll see in Section 6.

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 childcare use as inputs – jointly with the mother's work and childcare decision rules, thus implementing a selection correction. Her results suggest that one full year of maternal work and childcare 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 results generalize. Second, welfare rules have large effects on work/childcare use by single mothers, so as instruments they provide a strong basis for identification. It is difficult to find plausibly exogenous variables that impact behavior of married women so strongly.

Aside from the above studies, several papers estimate cognitive ability production functions, but for children old enough to be in pre-school or primary school (as opposed to childcare). Currie and Thomas (1995) look at pre-school inputs (i.e., Head Start), Liu, Mroz, and van der Klaauw (2003) study 5 to 15 year olds, and Todd and Wolpin (2007) and Cunha and Heckman (2008) look at 6 to 13 year olds. Thus, none of these studies address how childcare affects child outcomes.[7]

**2.2 Relationship between Test Scores and Subsequent Outcomes (Wages, Education, etc.)**

Several studies examine the relation between childhood test scores and subsequent outcomes like 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

---

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

[7] 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.

a significant positive relationship between scores at 7 and earnings at 23. Robertson and Symons (2003) and Harmon and Walker (1998) find a positive association between scores at age 7 and earnings at 33. Hutchinson, Prosser, and Wedge (1979) and Currie and Thomas (2001) find test scores at 7 are good predictors of scores at 16.

But these prior studies look at tests at age 7 or older,[8] while we study tests at ages 3 to 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 scores at age 4 and PIAT scores at ages 5 to 6 are significantly related to with educational attainment. For example, consider a one-point increase in the math score at age 6 (roughly a 1% increase, as the mean is 99.7). Holding background variables like mother's AFQT and education fixed, this is associated with an increase in education (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, as our estimates imply that a year of informal childcare reduces test scores by 2.6%. This translates into a drop in completed schooling of .050 to .065 years, a large effect.[9] Also, a striking aspect of the results is that mother's AFQT is <u>not</u> 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 (see Web Appendix A for further details).

## 3. Construction of Instruments using Welfare Rules and other Policy Variables

To deal with endogeneity of maternal work/childcare, 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 their labor supply (see Moffitt (1992)). To construct our instruments, we collected information on State welfare policies from many sources (see Fang and Keane (2004) or 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).

### 3.1. Benefit Termination Time Limits

Under AFDC, single mothers with children under 18 could receive benefits as long as they met income and asset screens. But under Section 1115 Waivers and TANF, States could set time limits on benefits. There is substantial heterogeneity across States in how the limits were set. For

[8] The studies we are aware of that use U.S. data (i.e., Neal and Johnson (1996), Murnane, Willet, and Levy (1995) and Zax and Rees (1998)) look at test scores measured at age 14 or later (See Web Appendix A for details).
[9] The following back-of-the-envelop 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.

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, e.g., 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 ($TL\_HIT_{ist}$), and her maximum potential remaining eligibility ($REMAIN\_TL\_ELIG_{\text{ist}}$).

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 she could have hit the limit, but only if her oldest child is at least 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, as 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).

### 3.2. Work Requirement Time Limits and Work Requirement Exemptions

TANF recipients must commence "work activities" within two 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*.

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

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 ($BEN(1)_{st}$ and $BEN(2)_{st}$), the "benefit reduction rate," which is the rate benefits are reduced with earnings ($PERC\_DISREGARD_{st}$), and the "earnings disregard," a part of earnings that is disregarded before calculating any benefit reduction ($FLAT\_DISREGARD_{st}$). Note that in constructing grant levels we do not condition on actual family size, as we treat fertility as endogenous. Grant levels have always differed greatly by State, but TANF created substantial State heterogeneity in disregards as well.

### 3.4. 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 locate absent parents and establish paternity. CSE expenditures increased from $2.9 billion in 1996 to $5.1 billion in 2002, increasing the likelihood that a single woman can collect child support. We measure State level CSE activity by dividing State CSE expenditures by the State population of single mothers ($ENFORCE_{st}$). The CCDF helps States provide subsidized childcare for low-income families. But States can design their own programs, so a great deal of heterogeneity has emerged. We use the State CCDF expenditure per single mother ($CHILDCARE_{st}$) to measure the availability and generosity of childcare subsidies in a State.

**3.5. 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 1994-96. The EITC subsidy rate varies by family size.[10] We use as instruments the wage subsidy rates for families with one and 2+ children ($EITC(1)_{st}$ and $EITC(2)_{st}$ respectively), using Federal and State EITC rules.[11] As with benefit levels (see Section 3.3) 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.*

**4. 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:

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

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

Several difficult issues arise in estimation of (2). A completely general specification, where the inputs and the ability endowment $\omega_i$ have different effects at each age, is infeasible due to

[10] For example, in 2003, the subsidy rates for families with one and two children were 34% and 40%, respectively.
[11] As of 2003, 17 States had enacted State earned income tax credits that supplement the federal credit.

proliferation of parameters.[12] Thus, we must restrict how inputs enter (2). Two simplifications, familiar from the human capital literature, are to assume: (i) only cumulative inputs matter, and (ii) the ability endowment has a constant effect over time. Letting $\widehat{X}_{it} = \sum_{\tau=1,t} X_{i\tau}$ be the cumulative input of $X$ up through $t$, and assuming inputs affect $ln\ A_{it}$ linearly, we obtain:[13]

(3) $$\ln A_{it} = \alpha_0 + \alpha_1 \widehat{T}_{it} + \alpha_2 \widehat{C}_{it} + \alpha_3 \ln \widehat{G}_{it} + \omega_i$$

Next, we further assume that the ability endowment $\omega_i$ is given by the equation:

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

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 measurement of maternal time. This can take various forms (e.g., "quality" time vs. the child watching TV while she does housework), but we don't observe these distinctions in the data. Thus, we distinguish only two types of time, time with the mother vs. time in childcare, and assume that $T_{it} = T\text{-}C_{it}$, where $T$ is total time in a period. Then, we can rewrite (3) as:

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

Thus, we can only estimate $\alpha_2 - \alpha_1$, the effect of time in childcare *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, etc.. 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, as well as preferences), the child's unobserved ability $\hat{\omega}_i$ ,[14] and child age. Specifically, we write:

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

where the stochastic term $\varepsilon_i^g$ captures the household's tastes for investment in the form of goods.[15] Note that prices of goods and childcare are not included in (6) as we cannot measure them, so in effect we assume they are fixed. One reason the prior literature failed to find good instruments for childcare

---

[12] 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 20·21/2=210 parameters for that input alone.
[13] It is convenient to let the cumulative goods input enter in log form, for reasons that will become apparent below.
[14] 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).
[15] This would arise due to heterogeneous preferences for child quality. $\varepsilon_i^g$ may also be influenced by the child's tastes.

is lack of sufficient price variation in the data.[16] Substituting (6) and (4) into (5) we obtain:

$$\begin{aligned}\ln A_{it} &= \alpha_o + (\alpha_1 T)\cdot t + (\alpha_2 - \alpha_1)\hat{C}_{it} + \alpha_3[\gamma_o + \gamma_1 Z_i + \gamma_2 \hat{\omega}_i + \gamma_3 \ln \hat{I}_{it} + \gamma_4 t + \varepsilon_i^g] + \hat{\omega}_i \\ &= (\alpha_o + \alpha_3\gamma_o) + (\alpha_1 T + \alpha_3\gamma_4)\cdot t + (\alpha_2 - \alpha_1)\hat{C}_{it} + \alpha_3\gamma_3 \ln \hat{I}_{it} + \alpha_3\gamma_1 Z_i + (1+\alpha_3\gamma_2)\hat{\omega}_i + \alpha_3\varepsilon_i^g \\ &= \varphi_o + \varphi_1 \cdot t + \varphi_2 \hat{C}_{it} + \varphi_3 \ln \hat{I}_{it} + \varphi_4 Z_i + \hat{\hat{\omega}}_i + \hat{\varepsilon}_i^g\end{aligned} \tag{7}$$

Equation (7) is estimable, as all independent variables are observed. However, endogeneity problems arise if the inputs $\hat{C}_{it}$ and $\ln \hat{I}_{it}$ 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 childcare may well be correlated with the child's ability endowment, and the work decision affects income as well. Furthermore, tastes for childcare may well be correlated with tastes for goods investment in children.[17] Thus, estimation of (7) by OLS is unlikely to be appropriate.

In order for welfare rules and local demand condition to be valid instruments for cumulative childcare and income in (7), they must be uncorrelated with $\hat{\omega}_i$ and $\hat{\varepsilon}_i^g$. This seems like a plausible exogeneity assumption.[18] These instruments are also likely to be powerful, as it is well known that welfare rules have important effects on work (and hence childcare) 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 childcare time relative to maternal time. However, suppose demand for goods also depends on childcare 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 childcare would be contaminated by the effect of any change in goods inputs that she may choose as a result of using childcare (holding income fixed). Web Appendix B 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.

---

[16] 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 $\gamma_4$. 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 ($G_{it}=\exp(\gamma_0)\cdot I_{it}$).
[17] 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 childcare usage in a negative direction, since not only the maternal time input but also the goods input is lower for children in childcare.
[18] One might be concerned that States that adopted stricter welfare reform tended to have relatively high/low test scores initially. But in Web Appendix G we present evidence that this was not the case.

Finally, we do not observe actual cognitive ability of children, but 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 measurement process:

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

where $d_{1t}$ and $d_{2t}$ are cognitive ability test dummies, which capture the mean differences in scores across the three tests,[19] and $\varepsilon_{it}$ combines measurement error and shocks to the development path. A key question is whether it is appropriate to pool the three tests. Below we provide evidence that it is, provided 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:

(9) $$\ln S_{it} = \varphi_o + \varphi_1 \cdot t + \varphi_2 \widehat{C}_{it} + \varphi_3 \ln \widehat{I}_{it} + \varphi_4 Z_i + \eta_1 d_{i1t} + \eta_2 d_{i2t} + \eta_3 d_{it} X_i + \upsilon_{it}$$

where $\upsilon_{it} = \hat{\hat{\omega}}_i + \hat{\varepsilon}_i^g + \varepsilon_{it}$. 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 childcare use and observed characteristics of the mother (such as education and AFQT). And we test for differences in the effect of childcare depending on characteristics of the provider (i.e., formal vs. informal) and of the child (i.e., age, race, gender).

## 5. Data

### 5.1. Construction of the Sample

We use data from the National Longitudinal Survey of Youth 1979 cohort (NLSY79). Panel members were 14-21 on Jan. 1, 1979, and 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/childcare decisions were greatly affected by policy changes in our sample period.

We require that women in our sample be single (or not cohabitating with a male) during five 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

[19] 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.

may have multiple children, which may affect resources allocated to any one child.[20] Thus, we include the number of children in $Z_i$ in (9). Of course, number of children may be endogenous in (9), e.g., if there is a quality/quantity tradeoff, 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 impacted 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 pre-reform period some of our instruments, like AFDC grant levels and local demand conditions, varied greatly across States and over time, also providing an important source of identification. And, in the post-reform period, some States adopted "strict" vs. "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 non-white, and less likely to work during the first year after childbirth (39% vs. 47%).

**5.2. 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 five years). We also use retrospective data gathered in 1986, 1988, 1992, and 1994-2000 to construct quarterly childcare usage histories for the first three years of a child's life.

Unfortunately, the NLSY does not report hours of childcare. It contains only an indicator for whether the mother used childcare for at least 10 hours per week. This is inadequate to determine if care was full or part-time. However, by combining the childcare variable with data on work, we can make a reasonable imputation. It is particularly helpful that, for single mothers, we know that childcare hours must be at least as great as their work hours.[21] Thus we use the following procedure:

If a woman reports using 10+ hours per week of childcare, we assume she used childcare during the quarter. If she worked full-time (i.e., 375+ hours in a quarter), we assume childcare must have been full-time, which seems clear. However, if the mother did not work (i.e., <75 hours in a

[20] 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 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.

[21] Note that we define "childcare" as all non-maternal 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 childcare.

quarter) but still reported using childcare – not a common state for single mothers – it seems likely childcare 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 childcare was part-time in this case. We admit this assignment is not so obvious. However, we experimented with assigning full-time childcare instead, and found it had almost no effect on the results.[22] Thus, we define the function:

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

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

As we noted earlier, complete childcare histories are only available for three years after childbirth. We impute childcare choices in years 4 and 5 based on current work and work/childcare histories. Again, imputation is made easier by the fact that childcare and work are so closely linked for single mothers. Thus, we set $I_t^c$ =1 or 0.5 for mothers who work full 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 childcare choice using a probit model we estimate using observed work and childcare histories. As the probit coefficients in Table B1 in appendix B show, childcare use by non-working mothers is very well predicted by lagged childcare and work choices. For quarters when we observe childcare (i.e., the first 12), our imputations from this probit are correct in 88.7% of cases.[23]

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 Peabody Picture Vocabulary Test (PPVT) at ages 3, 4 and 5, and the Peabody Individual Achievement Test (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.

[22] This is not surprising as the cumulated childcare variables constructed in these two ways have a correlation of .98.
[23] 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 a woman who used childcare at $t$-1 continues to do so at $t$ is 92.4%. Indeed, the unconditional persistence rate in childcare use is 93.5%, while that in non-use is 89.1%.

## 5.3. 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). 64% of the mothers worked prior to giving birth at an average hourly wage of $9.14 in 2007 dollars. Average annual household income is $22.7 thousand (2007) dollars. During the 20 quarters after childbirth mothers use 0.355 units of childcare per quarter, for a total of 7.1 quarters, on average. Comparing the '79-'93 and post-'93 periods, the childcare usage rate increases 10 points (from 59% to 69%).

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

Table C1 in Appendix C contains descriptive statistics for test scores. Mean scores on the PPVT, PIAT-M and PIAT-R are roughly 80, 95 and 101.[24] Note there is no clear age pattern in scores, as they are age adjusted. Interestingly, score differentials between children who are white vs. non-white and who have high-school graduate vs. high-school drop out parents are already apparent in the PPVT at age 3, *and there is no discernable pattern of these differentials growing over time*. This highlights the importance of studying determinants of test scores at very early ages.

# 6. Estimation Results

## 6.1. The Reduced Form Regressions for the Endogenous Variables

The first stage of 2SLS, and 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 childcare, income since birth, and number of children). The procedure is complicated by the fact that the instruments are time varying, and 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 (1) of Appendix 4.

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.[25] For cumulative childcare these are .1735 and .1483, respectively. Column (4) shows the incremental $R^2$s from adding the excluded instruments. For cumulative childcare, this is .0908, and the (cluster robust) F-test for joint significance of the excluded instruments is 14.74 (the 1% critical value is

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

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

1.47). These results suggest that our instruments are reasonably powerful, especially compared to those used in earlier attempts to study effects of maternal employment (see Section 2.1).

We omit the reduced form regressions, to conserve on space.[26] But it is worth noting that the 78 policy instruments have reasonable coefficients. The strongest predictors of cumulative childcare are: (i) if a State had a work requirement, which has a strong positive effect on work/childcare use, and a t-stat of 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 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 childcare) 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 childcare) 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.

### 6.2. Baseline Specification of the Child Cognitive Ability Production Function

Before arriving at baseline specification, we first test if it is appropriate to pool data from the three tests. Pooling is desirable if appropriate, as it leads to an efficiency gain. Web Appendix C 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* we allow for race/test and AFQT/test interactions (i.e., let $X_i$ in (9) include race and AFQT). Specifically, we find that non-whites 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.

[26] 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 equation (9) contains 3 endogenous variables, there are 75 over-identifying restrictions.

Next, we assess whether a cumulative or current childcare specification is most appropriate. Web Appendix D reports estimates of both models. Starting from the current childcare specification, if we add four years of lagged childcare the p-value for their joint significance is .0329. Thus, we reject that only current childcare matters. Also, a $\chi^2$ test for equality of coefficients on current and lagged childcare 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 childcare is essentially zero and insignificant. But the 2SLS estimate based on our 78 excluded instruments is -0.36% per quarter (-1.4% per year), and 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 towards 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 childcare 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 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 childcare 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.[27]

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 the last column 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.[28] But of course mother's education and AFQT may also be

[27] Repeating the calculation at the end of Section 2, this gives about a 6% decrease in the number who attend college.
[28] This is reminiscent of findings by Keane and Wolpin (2001) and Cameron and Heckman (1998) 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.

important for other reasons: genetic transmission of ability, more educated mothers may be better at teaching children, etc.. We do not attempt to disentangle these alternative mechanisms.

**6.3. Comparison of Alternative Estimation Methods**

As we noted earlier, we use the LIML estimator of Anderson and Rubin (1949) because both theory and Monte Carlo evidence suggest it is less subject to "many instrument" bias than 2SLS. Our results in Table 6 appear consistent with this, as the 2SLS estimate of the childcare effect, using 78 over-identifying instruments, is shifted about 25% of the way towards 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 3 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. And 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 instrument 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 childcare, 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 unemployment. Factor 19 captures work requirements, sanctions and strictness of welfare rules. Factor 21 captures prior work experience interacted with child age.[29] Factor 24 captures CCDF spending, CSE and EITC. Factor 26 captures benefit levels, remaining eligibility, EITC and local wages.

Table 7B examines how well the 14 factor instruments predict cumulative childcare. 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 childcare use as expected. Factor 12 also has a substantial negative coefficient (t=-5.5), implying higher unemployment reduces work and childcare. And factor 21 has a positive coefficient (t=7.6), so work experience/cumulative childcare grows more quickly with child age for mothers with more prior experience. [Note that factors with the most explanatory power for childcare do not correspond to those that had the largest eigenvalues in the factor analysis].

The Shea (squared) partial correlations for childcare, 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 childcare) 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 important for both income and children (but not childcare). 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 7A 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 childcare is hardly affected by using the

[29] Note: 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 discussion prior to equation (3)). The interaction is useful for predicting cumulative childcare as mothers who work more prior to childbirth tend to work more afterward.

14 factors in place of the full set of 78 instruments. Second, and most importantly, using the factors as instruments leads to an increase in efficiency. The standard error of the LIML estimate of the childcare 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 towards OLS, vs. 25% when using all 78 instruments. Indeed, the Cragg-Donald weak instrument test statistic, which was 5.80 when we used all 78 instruments, improves to 15.33 using the 14 factors. In the case of 3 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.

The last column of Table 7A reports results using the first 21 principle factors as instruments. As we see in Table 7B, 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 Web Appendix E for further discussion). However, it is comforting this approach leads to similar results. That is, is gives just a slightly larger estimate of the childcare effect (-0.63% per quarter, t=-2.3).

**6.4. Robustness of the Results with Respect to Specification of the Main Equation**

In Table 8 we consider 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 20 and over 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 instruments are correlated with mothers' age at childbirth, due to the timing of waivers/TANF. Waivers were first implemented in some States in '92-'93 and, as Fang and Keane (2004) note, binding work requirements first hit significant numbers of women in 1995-6. Thus, for a child born prior to 1990, it is unlikely waivers could have influenced the mother's labor supply before the child was 6. In the NLSY79, 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 20s and 30s, while prior to 1990, many were to teenage mothers. So, loosely speaking, stricter welfare rules and greater childcare usage will be *positively* correlated with maternal age at childbirth. Then, if (i) mother's age at childbirth has a positive effect on child cognitive ability, and (ii) 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/childcare use on child cognitive ability test scores.

But results in the columns (1)-(2) of Table 8 suggest this is not a concern. Conditional on measures of her 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 maternal youth is beneficial for child outcomes (the under 20 dummy is positive in column (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. Web Appendix F contains more evidence on this issue.

Next, in Table 8 column (3) we consider the impact of dropping the mother's AFQT score. This leads to a slight increase in the childcare 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; but AFQT is a better proxy, so when it is included the income variable drops out.[30]

Third, we consider 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 childcare effect. Thus, column (4) includes separate controls for number of children aged 0-5 and 6-17 (both treated as endogenous). The estimates imply that young siblings (0-5) have a larger negative effect on test scores. However, this has little impact on the estimated childcare 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 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 childcare 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, as it deals with any (cross-State) correlation between the instruments and *unobserved*

[30] 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%. But the model still includes such variables as mother's education and pre-childbirth wage, which also proxy for her permanent income/skill endowment.

child ability endowments (e.g., if States with relatively low ability children adopted stricter welfare reform it would bias the childcare effect negatively). Adding State fixed effects to the main equation actually shifts our estimate of the cumulative childcare effect from -.53% to a much larger negative value of -.98%. But 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 the note to column (6)), 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 – e.g., if States with low ability children adopted stricter welfare rules. But in Web Appendix G we show there is no significant difference in average child test scores (in the pre-reform period) between States that adopted more vs. 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.[31] And 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.[32] Given these considerations, we do not adopt fixed effects as the baseline model.

Finally, in Table 8 column (7) we consider a model where test scores enter in levels. The estimates imply that a quarter of childcare reduces scores by 0.5 points. As the mean score is 91.9, this corresponds to -0.54% per quarter, almost identical to the effect in the log specification.

**6.5. 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 robustness of our results to the instrument list. Table 9 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.

In column (2) we exclude CCDF spending. CCDF shifts the effective price of childcare, so it

[31] 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 Section 2].

[32] 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 childcare 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.

arguably belongs in equation (6). But excluding this instrument has little impact on the estimated childcare effect. In column (3) we use only the main features of TANF as instruments: time limits, work requirements and disregards. This increases the estimated childcare 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 childcare 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 90s (e.g., TANF, CCDF, EITC), using only instruments that varied across States/time regardless. These are State welfare grant levels and local demand conditions. Here the childcare 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/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. It has little impact on the estimated childcare effect.

Recall that some of our instruments are tailored to individuals based on ages of their children (see Section 3.1). 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 childcare 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 944 mothers in our sample have multiple children. But in Web Appendix H we show this makes little difference. Within cluster correlations are quite small at these levels. Clustering by child increases standard errors by 25% to 40%, depending on the instrument set.

In summary, our result of a negative childcare effect is robust to a wide range of alternative instrument sets, with estimates ranging from –1.7% to –3.0% per year, and 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 over-identifying restrictions for <u>any</u> instrument set we consider.

### 6.6. Heterogeneity in the Effect of Childcare by Type of Mother and Child

In Table 10 we assess how the effects of childcare vary with characteristics of the mother or child. In the first three columns we include interactions between cumulative childcare 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 childcare for a typical mother.

In Table 10 column (1) the interaction between mother's education and childcare 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, e.g., that if a mother's education is 4 years above the sample average, then the negative childcare effect goes from -.46% to -.81%. The later estimate has a standard error of .28, and hence a t-stat of -2.90. Thus, we have clear evidence that childcare 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 childcare and number of children in the household is very small and insignificant, implying the effect of childcare does not depend on number of siblings.

Column (4) includes an interaction of childcare with gender. It is significant, and the point estimates imply the effect of childcare 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 childcare to vary linearly with age.[33] The age/childcare interaction variable is insignificant, so we find no evidence of differences by child age.

**6.7. The Effects of Different Types of Child Care**

So far we have reported on effects childcare *in general*, but it seems likely that the type of care matters. That is, *formal* center-based care by trained providers (e.g., daycare centers, pre-school) may have different effects from *informal* care provided by relatives (e.g., grandparents, siblings) or non-relatives. Thus, we estimated versions of equation (9) in which effects of childcare are allowed to vary by type of care.[34] First, Table 5 Panel B present results of the reduced form regressions for childcare inputs of different types (i.e., formal, informal, etc.). Even at this more refined level, the welfare policy/demand condition variables are reasonably powerful instruments. The marginal $R^2$s 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 the excluded instruments are highly significant.

---

[33] We let the impact of childcare be a linear function of child age, as in $\sum_{\tau=1,t} (\varphi_{20} + \varphi_{21} \cdot t) C_t$. This expression can be rewritten $\varphi_{20} \widehat{C}_t + \varphi_{21} \cdot \sum_{\tau=1,t} (t \cdot C_t)$, where $\sum_{\tau=1,t} (t \cdot C_t)$ is the age/childcare interaction variable referred to in the text. We add this variable to equation (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).

[34] Recall that we do not have direct measures of childcare use in years 4 and 5, and we impute this using the procedure described in Section 5.2 and in Appendix B. Having imputed childcare 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 childcare for two consecutive periods, the own transition rates for formal and informal care are both roughly 98%.

The reduced form regression coefficients (not reported) also appear 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 vs. 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 childcare in general, but that of using informal care is reduced much more. Education interactions imply welfare rules have less influence on more educated women.[35]

Strikingly, the LIML results in Table 11 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 childcare causes a 2.6% reduction in test scores. The estimated effect for formal care is actually positive, but 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 U.S. 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 childcare in general. The results are reported in Web Appendix I, and they show that the result is quite robust to changes in specification, the instruments, etc..[36]

In Table 11 column (5) we divide informal childcare into that provided by relatives (most often grandparents) vs. non-relatives (e.g., family daycare). 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 non-relatives 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 childcare 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 Millenium 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

[35] AFQT effects are subtler, but 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 only for relatively high AFQT women (consistent with high AFQT women being more likely to find work).

[36] 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.

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, as 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 (Millwood (1998), Goodfellow and Laverty (2003)). Prior literature also suggests center-based care has two advantages over informal care: (1) trained care providers may provide more cognitive simulation to children than informal providers,[37] and (2) center-based care may provide more stimulating interaction with other children and more educational activity than informal care.[38]

In the last column of Table 11 we interact type of childcare 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 center-based care is beneficial for low socioeconomic status (SES) children (see, e.g., Currie and Thomas (1995) on Head Start, or Pungello, Campbell, and Barnett (2006) on the Perry Pre-School and Abecedarian experiments).

### 6.8. The Effect of Welfare Reform on Child Test Scores

Our focus has been on using welfare policy changes of the mid-90s as a source of exogenous variation to help identify effects of childcare 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 childcare effect, we felt 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-

[37] 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.

[38] Ideally, we would also like to examine how effects of childcare differ by direct measures of childcare quality. But it is the NLSY lacks good quality measures. Hence, we have instead differentiated between formal and informal care. But the notion that formal care is superior is consistent with the evidence on who uses it. In Web Appendix J we present a logit for whether a mother uses formal or informal care (conditional on childcare 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, as it is typically used by women who can afford more expensive care. Similarly, Web Appendix J also presents a logit for use of relatives vs. non-relatives (conditional on using informal care). The more educated, urban women with fewer children are more likely to use non-relatives, suggesting that non-relatives provide higher quality care than relatives.

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).

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 childcare. As we see in Table 4, the childcare 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 childcare 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%.

**6.9. Childcare and Non-Cognitive Outcomes**

Of course, childcare may also affect non-cognitive outcomes. Indeed, it is possible that any negative effects on cognitive outcomes could be outweighed positive effects on non-cognitive outcomes. Table 12 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 childcare on the incidence of behavioral problems. However, the LIML results suggest a quantitatively large adverse effect of +0.46% per quarter of childcare.[39] But 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=1730) 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 non-cognitive skill.

[39] We use 19 factors chosen by the same procedure described in Sect. 6.3. Because the sample here is different (N=1730) 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.

## 7. Conclusions

In this paper we have used the children of single mothers in the NLSY79 to assess the impact of childcare use on child cognitive achievement measured at ages 3 to 6. To deal with endogeneity of childcare, we utilize the (plausibly) exogenous variation in work/childcare choices of 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 childcare. This event provides a good opportunity to extend the literature on effects of childcare on child outcomes – a literature which has been limited by the difficulty of finding plausible instruments for childcare use.

Our main results indicate that the effect of childcare on children's achievement is negative, significant and rather sizeable. Estimates of our baseline model imply that one year of full-time childcare 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 childcare, types of children, and types of mothers. What drives the negative estimate of the childcare effect is that most (i.e., about 75%) childcare used by single mothers is informal (i.e., care by grandparents or other relatives, or by non-relatives in non-center based settings). Our estimates imply that a year of informal childcare 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 childcare 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., 4, 5 and 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 like 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 childcare translates into roughly a .050 to .065 year reduction in completed schooling.

**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: $TL\_LENGTH_{st}$-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{$AGE\_OLDEST\_CHILD_{ist}$, $ELAPSED\_WR_{st}$} & $AGE\_YOUNGEST\_CHILD_{ist} \geq AGE\_CHILD\_EXEM_{st}$] |
| $ELAPSED\_WR\_HIT_{ist}$ | 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 non-compliance of work requirement in state $s$ at time $t$. |
| $EXEMP_{st}$ | Number of work requirement exemptions in state $s$ |
| | AFDC/TANF Benefits (BEN) [a] |
| $BEN(1)_{st}$ | Real AFDC/TANF maximum benefits for a family with 1 child |
| $BEN(2)_{st}$ | Real AFDC/TANF maximum benefits for a family with 2 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 1 child |
| $EITC(2)_{st}$ | EITC phase in rate for a family with 2 children |
| $CHILDCARE_{st}$ | CCDF expenditure per single mother in state $s$ at time $t$. |
| $ENFORCE_{st}$ | 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 $t$. |

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.

[a] Benefits are put in real terms using metropolitan area and regional CPI indices computed by the Bureau of Labor Statistics, given the SMSA and State of residence of each woman in the sample.

**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 |
| $I[AGE_i>=33]$ | Dummy for whether mother is older than 32 years old |
| $EDUC_i$ | Mother's educational attainment at childbirth |
| $AFQT_i$ | Mother's AFQT score |
| $AFQT_i*dPPVT$ | Mother's AFQT score interacted with a dummy for PPVT test |
| $I[WORK\_BEF]_i$ | Dummy for whether mother worked prior to childbirth |
| $I[WORK\_BEF]_i$ x $SKILL_i$ | Work dummy interacted with mother's skill* |
| $EXPBEF_i$ | Mother's total work experience (in number of years) prior to childbirth |
| $EXPBEF_i *age_i$ | EXPBEF interacted with mother's age |
| $MARAFT_i$ | Mother's marital status at time of child's test |
| $URBAN_i$ | 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) |
| $RACE_i*dPPVT$ | Child's race interacted with a dummy for PPVT test |
| $RACE_i*dMATH$ | 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 |
| $NUMCHILD_{0\text{-}5\ i}$ | Number of children 0-5 years of age |
| $NUMCHILD_{6\text{-}17\ i}$ | Number of children 6-17 years of age |
| $C_{it} *EDUC_i$ | Cumulative child care use interacted with mother's education |
| $C_{it} *AFQT_i$ | Cumulative child care use interacted with mother's AFQT score |
| $C_{it} *NUMCHILD_i$ | Cumulative child care use interacted with number of children |
| $C_{it} *NON\text{-}WHITE_i$ | Cumulative child care use interacted with non-white dummy |
| $C_{it} *BLACK_i$ | Cumulative child care use interacted with black dummy |
| $C_{it} *HISP_i$ | Cumulative child care use interacted with hispanic dummy |
| $TIME$ | Calendar time trend (0 to 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.

**Table 3**
**Mean Characteristics of Mothers in the Sample**

| Description | All mothers in NLSY | Single mothers at childbirth only | Single mothers for 5 yrs 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 | 0.47 | 0.73 | 0.79 | 0.83 |
| | (0.499) | (0.445) | (0.404) | (0.379) |
| Hourly wage before childbirth (first child)[&] | 6.32 | 4.74 | 4.90 | 4.39 |
| | (7.71) | (8.23) | (9.85) | (2.01) |
| Total number of children of mother | 2.9 | 3.1 | 3.1 | 3.1 |
| | (1.37) | (1.57) | (1.61) | (1.53) |
| Father present at birth | 0.55 | - | - | - |
| | (0.497) | | | |
| Observations | 4,814 | 2,528 | 1,820 | 1,464 |
| Cases with wages at childbirth observed | 2,622 | 1,208 | 753 | 670 |

Our sample screens are (1) The mother does not have a husband/partner for 5 years after childbirth and (2) The child has at least one test score observation.

Standard deviations in parenthesis.

[&]1983$. These correspond to $13.15, $9.86, $10.20 and $9.14 respectively in 2007$.

**Table 4**
**Summary of Variables used in the Empirical Analysis**

| Variable | Mean (standard dev.) | Variable | Mean (standard dev.) |
|---|---|---|---|
| log(Test Score) | 4.49855<br>(0.1861)* | Urban | 0.8189<br>(0.3851) |
| Mother's education | 11.208<br>(1.8972) | Average yearly income (Thousands)[2] | 10.9274<br>(13.568) |
| Mother's age | 23.136<br>(4.5820) | Cumulative income (Thousands)[3] | 51.1787<br>(67.415) |
| Boys (Children of single mothers) | 0.4976<br>(0.5001) | Units of child care per quarter[4] | 0.3546<br>(0.3064) |
| Hispanic or Black | 0.8262<br>(0.3790) | Cumulative child care use (Quarters) | 7.0923<br>(6.1273) |
| Birthweight | 111.97<br>(21.976) | Labor participation rate (avg 1979-1993) | 48.23<br>(6.5) |
| Mother worked before giving birth | 0.6431<br>(0.4792) | Labor participation rate (avg 1994-1999) | 60.40<br>(5.4) |
| Wage rate prior to giving birth[1] | 4.3938<br>(2.0075) | Welfare participation rate (avg 1979-1993) | 58.93<br>(4.1) |
| Accumulated work experience prior to giving birth (number of years) | 4.7202<br>(6.0088) | Welfare participation rate (avg 1994-1999) | 44.65<br>(12.1) |
| Never married after childbirth | 0.7215<br>(0.4483) | Childcare use rate (avg 1979-1993)[5] | 59.05<br>(5.0) |
| Separated after childbirth | 0.1540<br>(0.3611) | Childcare use rate (avg 1994-1999) | 69.27<br>(5.9) |
| Divorced after childbirth | 0.1158<br>(0.3201) | | |

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

[1] 1983$ (values used in the regressions). This corresponds to $9.14 in 2007$.

[2] 1983$ (values used in the regressions). This corresponds to $22.7 thousand in 2007$.

[3] 1983$ (values used in the regressions). This corresponds to $106.5 thousand in 2007$.

[4] One quarter of full-time child care use is 1 unit and one quarter of part-time child care use is 1/2 unit.

[5] It is equal to 1 if child care is used (either full-time or part-time).

**Table 5**
**Explanatory Power of the Instruments**

| Input | Partial correlation squared | Shea partial correlation squared | $R^2$ with exogenous variables only | Incremental $R^2$ | F-statistic[#] | P-value |
|---|---|---|---|---|---|---|
| A. Endogenous variables in the baseline model | | | | | | |
| Cumulative Child Care Use | 0.1735 | 0.1483 | 0.4765 | 0.0908 | 14.740 | 0.0000 |
| Current Child Care Use | 0.1314 | 0.0986 | 0.3378 | 0.0807 | 10.970 | 0.0000 |
| Cumulative Income | 0.1112 | 0.1163 | 0.2163 | 0.0872 | 26.120 | 0.0000 |
| Current Income | 0.0873 | 0.0864 | 0.1130 | 0.0767 | 3.4200 | 0.0000 |
| Number of Children | 0.3944 | 0.3252 | 0.2469 | 0.2970 | 25.220 | 0.0000 |
| B. Other endogenous variables in additional models | | | | | | |
| Cumulative Formal Child Care | 0.0943 | 0.0997 | 0.0906 | 0.0858 | 51.470 | 0.0000 |
| Cumulative Informal Child Care | 0.1392 | 0.1446 | 0.3385 | 0.0921 | 16.310 | 0.0000 |
| Cumulative Child Care by Nonrelatives | 0.0956 | 0.1001 | 0.0893 | 0.0871 | 1.8100 | 0.0001 |
| Cumulative Child Care by Relatives | 0.1277 | 0.1461 | 0.2255 | 0.0989 | 14.360 | 0.0000 |
| Age Weighted Cumulative Child Care | 0.1892 | 0.1043 | 0.4556 | 0.1030 | 25.260 | 0.0000 |

Instruments are: variables in the main equation (see Table 2), mother´s age and 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.
[#] F-stat is cluster robust. Critical value at 1% is 1.47 (78 d.f. in the numerator and 1463 in the denominator).

**Table 6**
**Comparison of Results by Estimation Method**

Dependent Variable -> Log(Test Score)

| | OLS | GMM | 2SLS | FULLER[&] | LIML | LIML[a] |
|---|---|---|---|---|---|---|
| Cumulative Child Care | 0.00098 | -0.00389 | -0.00357 | -0.00521 | -0.00522 | -0.00533 |
| | (0.0008) | (0.0015) ** | (0.0021) * | (0.0028) * | (0.0028) * | (0.0025) ** |
| Log(Cumulative Income) | -0.00324 | 0.01142 | 0.00742 | 0.01035 | 0.01037 | 0.01062 |
| | (0.0057) | (0.0087) | (0.0164) | (0.0242) | (0.0242) | (0.0266) |
| Mother's education | 0.01051 | 0.01130 | 0.01214 | 0.01276 | 0.01276 | 0.01297 |
| | (0.0026) ** | (0.0023) ** | (0.0029) ** | (0.0032) ** | (0.0032) ** | (0.0030) ** |
| Mother's AFQT | 0.00059 | 0.00057 | 0.00063 | 0.00066 | 0.00066 | 0.00066 |
| | (0.0002) ** | (0.0002) ** | (0.0003) ** | (0.0003) ** | (0.0003) ** | (0.0003) ** |
| Number of Observations | 3,787 | 3,787 | 3,787 | 3,787 | 3,787 | 3,787 |
| R-squared | 0.3994 | 0.3891 | 0.3914 | 0.3847 | 0.3847 | 0.3844 |
| k [#] | | | | 1.039 | 1.040 | 1.005 |
| Weak/Many Instruments Test | | 5.80 | 5.80 | 5.80 | 5.80 | 15.33 |

In columns (2) to (5) we use as instruments the same 78 variables described in the footnote of Table 5. Additional control variables are listed in Table 2.

[a] Instruments are the 14 more relevant factors derived from the factor analysis of our original 78 instruments described in the footnote of Table 5.

Note: Residuals in the score equation depart modestly from normality, exhibiting some skewness and excess kurtosis. Neither consistency nor asymtotic 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.

[#] k is the parameter of the k-class estimator, which equals 1 for 2SLS and exceeds 1 for LIML.

[&] Fuller parameter alpha=1

** Significant at 5%; * Significant at 10%

**Table 7**
**A. Results Based on Factor Analysis of the Instruments**

Dependent Variable -> Log(Test Score)

| | LIML | | LIML | | FULLER[&] | | 2SLS | | GMM | | LIML | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 78 instruments | | 14 rotated factors | | | | | | | | factors[a] | |
| Cumulative Child Care | -0.00522 | | -0.00533 | | -0.00531 | | -0.00498 | | -0.00420 | | -0.00633 | |
| | (0.0028) | * | (0.0025) | ** | (0.0025) | ** | (0.0024) | ** | (0.0023) | * | (0.0027) | ** |
| Log(Cumulative Income) | 0.01037 | | 0.01062 | | 0.01060 | | 0.01021 | | 0.01061 | | -0.00094 | |
| | (0.0242) | | (0.0266) | | (0.0265) | | (0.0243) | | (0.0213) | | (0.0269) | |
| Mother's education | 0.01276 | | 0.01297 | | 0.01297 | | 0.01282 | | 0.01308 | | 0.01298 | |
| | (0.0032) | ** | (0.0030) | ** | (0.0030) | ** | (0.0030) | ** | (0.0029) | ** | (0.0033) | ** |
| Mother's AFQT | 0.00066 | | 0.00066 | | 0.00066 | | 0.00065 | | 0.00063 | | 0.00077 | |
| | (0.0003) | ** | (0.0003) | ** | (0.0003) | ** | (0.0003) | ** | (0.0003) | ** | (0.0003) | ** |
| Instrument Set | All | | ---------------------------14 rotated factors ----------------------------- | | | | | | | | Principal factors | |
| Number of Observations | 3,787 | | 3,787 | | 3,787 | | 3,787 | | 3,787 | | 3,787 | |
| R-squared | 0.3847 | | 0.3844 | | 0.3845 | | 0.3860 | | 0.3890 | | 0.3761 | |
| k [#] | 1.040 | | 1.005 | | 1.005 | | - | | - | | 1.006 | |
| Weak/Many Instruments Test | 5.80 | | 15.33 | | 15.33 | | 15.33 | | 15.33 | | 11.13 | |

In the first column we use as instruments the same 78 variables described in the footnote of Table 5. Additional control variables are listed in Table 2.
Instruments used in second through fifth columns are 14 factors derived from the factor analysis of our original 78 instruments described in footnote in Table 5.
[a] Instruments used in the last column are the first 21 (unrotated) factors.
Robust standard errors (Huber-White) by child clusters.
[&] Fuller parameter alpha=1.
[#] k is the parameter of the k-class estimator, which equals 1 for 2SLS and exceeds 1 for LIML.
** Significant at 5%; * Significant at 10%

**B. Explanatory Power of Instruments in First Stage Regressions for Childcare (Instruments in Table 7A)**

Dependent variable-> Cumulative Childcare

| Instruments listed in footnotes in Table 7A | Partial correlation squared | Shea partial correlation squared | Incremental $R^2$ | F-statistic | P-value |
|---|---|---|---|---|---|
| All 78 instruments | 0.1735 | 0.1483 | 0.0908 | 14.740 | 0.000 |
| 14 rotated factors | 0.1043 | 0.0967 | 0.0546 | 14.570 | 0.000 |
| 21 principal factors | 0.1039 | 0.0960 | 0.0544 | 12.630 | 0.000 |

$R^2$ of first stage regression with only exogenous variables=0.4765

**Table 8**
**Robustness with respect to the Specification of the Main Equation**

Dependent Variable -> Log(Test Score)

| | Baseline | Aditional mother's age controls | Removing AFQT | Children by ages | Year effects | State effects[&] | Test Score in levels[a] |
|---|---|---|---|---|---|---|---|
| Cumulative Child Care | -0.00533 | -0.00535 | -0.00613 | -0.00553 | -0.00490 | -0.00982 | -0.50258 |
| | (0.0025) ** | (0.0026) ** | (0.0031) ** | (0.0026) ** | (0.0026) * | (0.0053) * | (0.2118) ** |
| Log(Cumulative Income) | 0.01062 | 0.00259 | 0.07808 | 0.00564 | 0.00228 | 0.02863 | 1.20063 |
| | (0.0266) | (0.0279) | (0.0284) ** | (0.0280) | (0.0300) | (0.0344) | (2.2509) |
| Mother's education | 0.01297 | 0.01435 | 0.01355 | 0.01538 | 0.01329 | 0.01467 | 1.05541 |
| | (0.0030) ** | (0.0033) ** | (0.0036) ** | (0.0044) ** | (0.0031) ** | (0.0036) ** | (0.2532) ** |
| Mother's AFQT score | 0.00066 | 0.00070 | | 0.00067 | 0.00067 | 0.00069 | 0.07814 |
| | (0.0003) ** | (0.0003) ** | | (0.0003) ** | (0.0003) ** | (0.0003) ** | (0.0271) ** |
| Child's age | 0.03817 | 0.04054 | 0.01833 | 0.03812 | 0.04078 | 0.04225 | 3.38379 |
| | (0.0122) ** | (0.0126) ** | (0.0128) | (0.0124) ** | (0.0131) ** | (0.0131) ** | (1.0073) ** |
| Mother's age | | -0.01260 | -0.00101 | | | | |
| | | (0.0150) | (0.0169) | | | | |
| Mother's age squared | | 0.00024 | -0.00005 | | | | |
| | | (0.0003) | (0.0004) | | | | |
| I[age of mother$_i$<20] | 0.02368 | 0.00799 | 0.00811 | 0.04335 | 0.0089 | 0.02387 | 1.83958 |
| | (0.0116) ** | (0.0149) | (0.0157) | (0.0272) | (0.0115) | (0.0119) * | (0.9577) * |
| I[age of mother$_i$>=33] | 0.0060 | -0.0049 | 0.0032 | 0.0109 | 0.0003 | 0.0081 | 0.2064 |
| | (0.0256) | (0.0310) | (0.0347) | (0.0269) | (0.0257) | (0.0261) | (2.2426) |
| Mother's age at first birth | | -0.0017 | | | | | |
| | | (0.0019) | | | | | |
| Number of children | -0.02545 | -0.02726 | -0.02616 | | -0.0250 | -0.02975 | -2.15419 |
| | (0.0064) ** | (0.0069) ** | (0.0079) ** | | (0.0067) ** | (0.0080) ** | (0.5449) ** |
| Number of children 0-5 | | | | -0.0298 | | | |
| | | | | (0.0079) ** | | | |
| Number of children 6-17 | | | | -0.0095 | | | |
| | | | | (0.0205) | | | |
| Estimation Method | LIML | LIML | LIML | LIML | LIML | LIML | LIML |
| Number of Observations | 3,787 | 3787 | 3787 | 3787 | 3787 | 3787 | 3787 |
| R-squared | 0.3844 | 0.3842 | 0.3304 | 0.3808 | 0.3884 | 0.3646 | 0.3936 |
| k[#] | 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 |

Instruments are 14 factors derived from the factor analysis of our original 78 instruments described in the footnote in Table 5.

[a] The mean score is 91.9, so the point estimate implies a child care effect of -0.55 per quarter, or -2.2% per year, similar to the log results.

[#] k is the parameter of the k-class estimator, which equals 1 for 2SLS and exceeds 1 for LIML.

[&] Joint-significance Test for State F.E.: 30.25 (0.066)

Robust standard errors (Huber-White) by child clusters.

** Significant at 5%; * Significant at 10%

**Table 9**
**A. Robustness with respect to the Instrument List**

Dependent Variable -> Log(Test Score)

| | (1) Original set of IVs [&] | | (2) Excluding CCDF [a] | | (3) Only TL, WR and ED | | (4) Excludes TL, WR and ED | | (5) Only BEN and local demand | | (6) Original set without interactions | | (7) Only State-specific instruments [b] | | (8) Only State-specific IVs [b] without interactions | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cumulative Child Care | -0.00522 | | -0.00584 | | -0.00758 | | -0.00414 | | -0.00437 | | -0.00528 | | -0.00623 | | -0.00642 | |
| | (0.0028) | * | (0.0028) | ** | (0.0038) | ** | (0.0024) | * | (0.0024) | * | (0.0028) | * | (0.0025) | ** | (0.0026) | ** |
| Log(Cumulative Income) | 0.01037 | | 0.01026 | | 0.02118 | | 0.02054 | | -0.00294 | | -0.00445 | | -0.00300 | | 0.00529 | |
| | (0.0242) | | (0.0250) | | (0.0367) | | (0.0302) | | (0.0372) | | (0.0239) | | (0.0246) | | (0.0361) | |
| Mother's education | 0.01276 | | 0.01305 | | 0.01327 | | 0.01049 | | 0.01208 | | 0.01285 | | 0.01316 | | 0.01167 | |
| | (0.0032) | ** | (0.0032) | ** | (0.0034) | ** | (0.0036) | ** | (0.0036) | ** | (0.0030) | ** | (0.0034) | ** | (0.0037) | ** |
| Mother's AFQT | 0.00066 | | 0.00067 | | 0.00064 | | 0.00056 | | 0.00073 | | 0.00077 | | 0.00078 | | 0.00073 | |
| | (0.0003) | ** | (0.0003) | ** | (0.0003) | ** | (0.0003) | * | (0.0004) | ** | (0.0003) | ** | (0.0003) | ** | (0.0004) | * |
| Estimation Method | LIML | | LIML | | LIML | | LIML | | LIML | | LIML | | LIML | | LIML | |
| R-squared | 0.3847 | | 0.3816 | | 0.3705 | | 0.3843 | | 0.3860 | | 0.3820 | | 0.3767 | | 0.3720 | |
| 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 | 0.744 | | 0.713 | | 0.458 | | 0.979 | | 0.826 | | 0.892 | | 0.787 | | 0.911 | |
| Number of instruments | 78 | | 75 | | 58 | | 27 | | 18 | | 26 | | 63 | | 25 | |

[&] All 78 policy variables, local demand conditions and interactions described in footnote in Table 5. Unless noted in column heading, all specifications still include these interactions.

[a] See descriptions of instruments in Table 1. CCDF:Childcare Development Fund expenditures; TL:time limits; WR:work requirements; ED:earnings disregards; BEN:benefit amounts.

[b] 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).

Robust standard errors (Huber-White) by child clusters.

[#] k is the parameter of the k-class estimator, which equals 1 for 2SLS and exceeds 1 for LIML.

** Significant at 5%; * Significant at 10%

**B. Explanatory Power of Instruments in First Stage Regressions for Childcare (Instruments in Table 9A)**

Dependant variable-> Cumulative Childcare

| Instruments used in each column in Table 11A | Partial correlation squared | Shea partial correlation squared | Incremental $R^2$ | F-statistic | P-value |
|---|---|---|---|---|---|
| Original set of IVs | 0.1735 | 0.1483 | 0.0908 | 14.740 | 0.000 |
| Excluding CCDF | 0.1675 | 0.1426 | 0.0877 | 14.980 | 0.000 |
| Only TL, WR and ED | 0.1307 | 0.0944 | 0.0684 | 18.120 | 0.000 |
| Excludes TL, WR and ED | 0.1151 | 0.1094 | 0.0603 | 11.800 | 0.000 |
| Only BEN and local demand | 0.1046 | 0.0985 | 0.0548 | 16.210 | 0.000 |
| Original set without interactions | 0.1327 | 0.1072 | 0.0755 | 11.930 | 0.000 |
| Only State-specific IVs | 0.1443 | 0.1393 | 0.0756 | 9.550 | 0.000 |
| State-specific IVs without interactions | 0.1189 | 0.1144 | 0.0622 | 15.510 | 0.000 |

$R^2$ of first stage regression with only exogenous variables=0.4765

**Table 10**
**Heterogeneity in Effect of Maternal Time Inputs**

Dependent Variable -> Log(Score)

| | By maternal education | By maternal AFQT score | By number of children | By child's gender | By child's race | By child's race[a] | By child's age |
|---|---|---|---|---|---|---|---|
| Cumulative Child Care | -0.00456 | -0.00493 | -0.00502 | -0.00709 | -0.00814 | -0.00885 | 0.00465 |
| | (0.0029) | (0.0029) * | (0.0027) * | (0.0029) ** | (0.0041) ** | (0.0046) * | (0.0145) |
| $\widetilde{\text{Education}}$*(Cum. Child Care) | -0.00088 | | | | | | |
| | (0.0005) * | | | | | | |
| $\widetilde{\text{AFQT}}$*(Cum. Child Care) | | -0.00011 | | | | | |
| | | (0.00007) * | | | | | |
| $\widetilde{\text{(Number of Children)}}$*(Cum. Child Care) | | | -0.00008 | | | | |
| | | | (0.0013) | | | | |
| Male * (Cum. Child Care) | | | | 0.00331 | | | |
| | | | | (0.0015) ** | | | |
| Non-white * (Cum. Child Care) | | | | | 0.00371 | | |
| | | | | | (0.0030) | | |
| Black * (Cum. Child Care) | | | | | | 0.00469 | |
| | | | | | | (0.0031) | |
| Hispanic * (Cum. Child Care) | | | | | | 0.00325 | |
| | | | | | | (0.0036) | |
| Age weighted cumulative child care[&] | | | | | | | -0.00104 |
| | | | | | | | (0.0015) |
| Log(Cumulative Income) | 0.01379 | 0.01441 | 0.01004 | 0.00661 | 0.00827 | 0.00740 | 0.01089 |
| | (0.0244) | (0.0258) | (0.0244) | (0.0243) | (0.0245) | (0.0246) | (0.0242) |
| Mother's Education | 0.01858 | 0.01232 | 0.01288 | 0.01313 | 0.01278 | 0.01276 | 0.01251 |
| | (0.0042) ** | (0.0032) ** | (0.0031) ** | (0.0032) ** | (0.0031) ** | (0.0031) ** | (0.0032) ** |
| Mother's AFQT score | 0.00066 | 0.00171 | 0.00065 | 0.00068 | 0.00072 | 0.00076 | 0.00067 |
| | (0.0003) ** | (0.0007) ** | (0.0003) ** | (0.0003) ** | (0.0003) ** | (0.0003) ** | (0.0003) ** |
| No. of observations | 3,787 | 3,787 | 3,787 | 3,787 | 3,787 | 3,787 | 3,787 |
| Estimation Method | LIML | LIML | LIML | LIML | LIML | LIML | LIML |
| R-squared | 0.3868 | 0.3874 | 0.3861 | 0.3828 | 0.3867 | 0.3883 | 0.3831 |
| k [#] | 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 | |
| | | | | | | (0.2905) | |

$\widetilde{\text{Education}}$=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 footnote of Table 5 plus predicted childcare from the first stage interacted with mother's education (in column 1), with mother's AFQT (in column 2), with number of children (in column 3), with gender dummy (in column 4), with non-white dummy (in column 5), and with black dummy and hispanic dummy (in column 6).

Robust standard errors (Huber-White) by child clusters.

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

[&] Weights are the age of the child in number of quarters

[#] k is the parameter of the k-class estimator, which equals 1 for 2SLS and exceeds 1 for LIML.

** Significant at 5%; * Significant at 10%

**Table 11**
**Child Care Effects by Type of Care**

Dependent Variable -> Log(Score)

| | Mean (sd error) | Formal vs.Informal Child care | | Care provided by relatives vs. non-relatives | | Formal/Informal by education |
|---|---|---|---|---|---|---|
| | | OLS | LIML [&] | OLS | LIML | LIML |
| Cumulative INFORMAL child care | 5.8533 | 0.00067 | -0.00643 | | | -0.00638 |
| | (5.873) | (0.0008) | (0.0029) ** | | | (0.0031) ** |
| Relatives | 5.0077 | | | 0.00034 | -0.00751 | |
| | (5.736) | | | (0.0008) | (0.0029) ** | |
| Nonrelatives | 1.1454 | | | 0.00164 | 0.00683 | |
| | (3.355) | | | (0.0011) | (0.0069) | |
| Informal child care * I[mother college graduate] | | | | | | -0.00032 |
| | | | | | | (0.0023) |
| Cumulative FORMAL child care | 1.2229 | 0.00307 | 0.00302 | 0.00309 | 0.00472 | 0.00831 |
| (Daycare, Nursery, Pre-K, Other) | (3.055) | (0.0011) ** | (0.0066) | (0.0011) ** | (0.0066) | (0.0129) |
| Formal child care * I[mother college graduate] | | | | | | -0.00703 |
| | | | | | | (0.0101) |
| Log(Cumulative Income) | 3.6332 | -0.00343 | 0.00719 | -0.00331 | 0.00883 | 0.01094 |
| | (0.730) | (0.0057) | (0.0233) | (0.0057) | (0.0230) | (0.0233) |
| No. of observations | | 3787 | 3787 | 3787 | 3787 | 3787 |
| R-squared | | 0.4004 | 0.3776 | 0.4007 | 0.3520 | 0.3722 |
| k [#] | | | 1.0385 | | 1.0347 | 1.0379 |
| Weak/Many Instruments | | | 4.5 | | 4.48 | 3.3 |

Instruments are the same 78 variables described in the footnote of Table 5 plus predicted formal and informal childcare from the first stage interacted with a dummy variable for whether the mother is a college graduate or not (only in the last column).

[&] If State fixed effects are included in this regression they are not significant at the 10% level (see Web Appendix I Table I1). The coefficient on formal care remains small and insignificant, while that on informal care increases in magnitude to -.01013 (s.e. = .0061).

Robust standard errors (Huber-White) by child clusters.

[#] k is the parameter of the k-class estimator, which equals 1 for 2SLS and exceeds 1 for LIML.

**Table 12**
**Effect of Child Care on Non-Cognitive Outcomes**

Dependent Variable -> Log(Behavioral Problems Index)

| Dependent Variable -> | OLS | LIML[&] |
|---|---|---|
| Cumulative Child Care | -0.00083 | 0.00460 |
| | (0.0008) | (0.0032) |
| Log(Cumulative Income) | -0.00820 | 0.03270 |
| | (0.0057) | (0.0359) |
| Number of Children | -0.00448 | 0.01385 |
| | (0.0034) | (0.0070) ** |
| I[mother's age<20] | -0.01138 | -0.00523 |
| | (0.0099) | (0.0117) |
| I[mother's age>30] | -0.00508 | -0.01058 |
| | (0.0284) | (0.0334) |
| Mother's education | -0.00444 | -0.00587 |
| | (0.0027) * | (0.0035) * |
| Mother's AFQT score | -0.00005 | -0.00045 |
| | (0.0003) | (0.0004) |
| I[Non-white] | -0.03239 | -0.04504 |
| | (0.0119) ** | (0.0136) ** |
| Male | 0.03583 | 0.03699 |
| | (0.0075) ** | (0.0080) ** |
| Birthweight | -0.01446 | -0.01596 |
| | (0.0062) ** | (0.0066) ** |
| I[Work before] | -0.00152 | -0.01579 |
| | (0.0103) | (0.0149) |
| I[Work before] * skill | -0.00939 | -0.01950 |
| | (0.0093) | (0.0117) * |
| Experience before childbirth | 0.00256 | -0.00427 |
| | (0.0056) | (0.0077) |
| Experience * mother's age | -0.00012 | 0.00004 |
| | (0.0002) | (0.0002) |
| I[never married] | -0.09500 | -0.08624 |
| | (0.0301) ** | (0.0503) * |
| I[separated] | -0.09747 | -0.10221 |
| | (0.0310) ** | (0.0547) ** |
| I[divorced][a] | -0.10255 | -0.10005 |
| | (0.0324) ** | (0.0532) * |
| I[urban] | 0.00214 | 0.01572 |
| | (0.0108) | (0.0130) |
| Child's age | -0.00036 | -0.02148 |
| | (0.0069) | (0.0134) |
| Number of Observations | 1,730 | 1,730 |
| k [#] | | 1.018 |
| Weak/Many Instruments Test | | 5.08 |

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

[a] Excluded category: widowed.

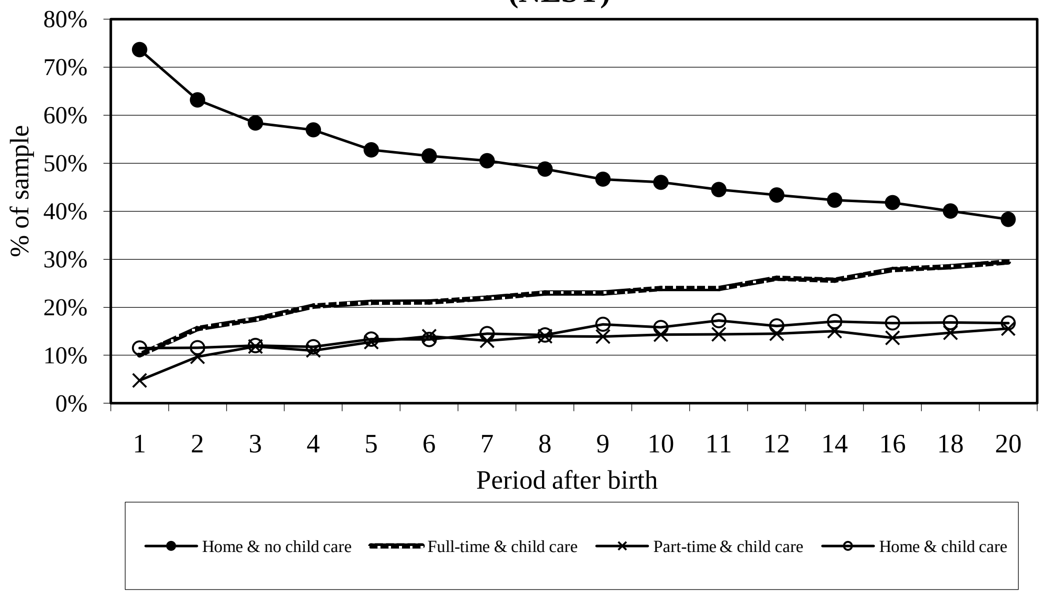
Robust standard errors (Huber-White) by child clusters.

[#] k is the parameter of the k-class estimator, which equals 1 for 2SLS and exceeds 1 for LIML.

** Significant at 5%; * Significant at 10%

**Figure 1**

**Employment and Childcare Choices of Single Mothers after Birth (NLSY)**




Note: The first 12 periods correspond to quarters (3-months) and the last four correspond to semesters.

**Appendix 1**
**Effect of early cognitive ability test scores on highest grade completed by 2000 (sample=young adults 18 years or older)**

Dependent Variable -> 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 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Test score | 0.00819 | 0.01574 | 0.00633 | 0.01627 | 0.00960 | 0.02092 | 0.01908 | 0.03166 | 0.02493 | 0.03744 |
| (Test and age in column heading) | (0.0041) ** | (0.0035) ** | (0.0046) | (0.0044) ** | (0.0048) ** | (0.0045) ** | (0.0049) ** | (0.0045) ** | (0.0056) ** | (0.0055) ** |
| Age of completed education measure# | 0.16449 | 0.06336 | 0.69629 | 0.68394 | 0.69097 | 0.66007 | 0.45305 | 0.41675 | 0.45629 | 0.40288 |
| | (0.1563) | (0.1575) | (0.0752) ** | (0.0717) ** | (0.0758) ** | (0.0723) ** | (0.0438) ** | (0.0409) ** | (0.0439) ** | (0.0411) ** |
| Highest grade completed by mother | 0.09231 | | 0.05216 | | 0.04901 | | 0.09646 | | 0.10179 | |
| | (0.0403) ** | | (0.0348) * | | (0.0343) | | (0.0270) ** | | (0.0268) ** | |
| Highest grade completed by father | 0.02147 | | 0.02069 | | 0.01948 | | 0.00833 | | 0.01065 | |
| | (0.0083) ** | | (0.0076) ** | | (0.0075) ** | | (0.0064) | | (0.0064) * | |
| Number of siblings | -0.14160 | | -0.14066 | | -0.12912 | | -0.08883 | | -0.08942 | |
| | (0.0586) ** | | (0.0543) * | | (0.0535) ** | | (0.0428) * | | (0.0424) ** | |
| Birthorder | -0.11146 | | -0.13111 | | -0.09435 | | -0.11223 | | -0.07853 | |
| | (0.0979) | | (0.0957) | | (0.0946) | | (0.0754) | | (0.0751) | |
| Race (1=Non-white) | 0.06958 | | 0.08739 | | 0.06939 | | -0.06182 | | -0.21639 | |
| | (0.1751) | | (0.1523) | | (0.1496) | | (0.1258) | | (0.1243) * | |
| Gender (1=Male) | -0.36024 | | -0.42114 | | -0.42716 | | -0.39478 | | -0.37505 | |
| | (0.1380) ** | | (0.1236) ** | | (0.1228) ** | | (0.1011) ** | | (0.1008) ** | |
| Mother's age at child's birth | -0.03878 | | -0.01219 | | -0.02523 | | 0.02586 | | 0.03390 | |
| | (0.0387) | | (0.0336) | | (0.0331) | | (0.0282) | | (0.0280) | |
| Mother's AFQT score | 0.00389 | | 0.00378 | | 0.00450 | | 0.00128 | | -0.00030 | |
| | (0.0038) | | (0.0033) | | (0.0033) | | (0.0029) | | (0.0028) | |
| Constant | 7.2531 | 8.3078 | -2.8778 | -3.8171 | -2.9770 | -4.0097 | -0.6925 | -0.1622 | -1.5869 | -0.6049 |
| | (3.1866) ** | (2.9599) ** | (1.8248) | (1.4088) ** | (1.7977) | (1.3892) ** | (1.2501) | -(0.1622) ** | (1.2644) | (0.9295) ** |
| No. of observations | 363 | 363 | 451 | 451 | 446 | 446 | 747 | 747 | 739 | 739 |
| Pseudo R-squared | 0.1578 | 0.0537 | 0.2791 | 0.2014 | 0.2912 | 0.2209 | 0.2365 | 0.1761 | 0.2457 | 0.1760 |

All estimated by OLS. ** indicates significance at 5% and * at 10%.
# The age of the young adult by 2000 if she is older than 18 years old. The average age is 21.8.
PPVT: Peabody Picture Vocabulary Test; PIAT: Peabody Individual Achievement Test

**Appendix 2**
**Probit to predict child care choices of non-working women in years 4 and 5 after childbirth**

| Dependent Variable-> Pr(using child care in *t* ) | | |
|---|---|---|
| Whether worked before giving birth | 0.5920 | |
| | (0.208) | ** |
| (Whether worked before) x (Avg. wage before) | -0.0642 | |
| | (0.040) | * |
| Total work experience (prior to giving birth) | -0.0060 | |
| | (0.019) | |
| Child's race | -0.0874 | |
| | (0.170) | |
| Child's gender | 0.0497 | |
| | (0.120) | |
| Mother's education | 0.0821 | |
| | (0.038) | ** |
| Total work experience since child birth | -0.3983 | |
| | (0.070) | ** |
| Total child care use since child birth | 0.2226 | |
| | (0.053) | ** |
| Whether used child care or not in *t-1* | 1.7801 | |
| | (0.164) | ** |
| Number of observations | 867 | |
| Pseudo-$R^2$ | 0.4585 | |

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

**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 0.5. Finally, if the mother does not work in a given period, we imputed a child care value of 0.5 if the predicted probability of child care use based on this model exceeds 0.65.We choose this threshold to obtain a smooth trend of child care use since childbirth and until the end of the fifth year.

**Appendix 3**
**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 |
| | (0.191) | (0.295) | (0.239) | (0.152) | (0.128) | (0.152) | (0.095) |
| 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) |
| Non-whites | 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 yrs+) | 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 yrs) | 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) |
| No. of observations | 339 | 512 | 438 | 598 | 663 | 584 | 653 |

PPVT: Peabody Picture Vocabulary Test
PIAT: Peabody Individual Achievement Test
Standard deviations in parenthesis.