

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

Raquel Bernal¹
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: July 31st, 2008

Abstract

We evaluate the effect of childcare vs. maternal time inputs on child cognitive development using the single mothers from the National Longitudinal Survey of Youth (NLSY79). To deal with non-random selection of children into childcare, we exploit the (plausibly) 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 (along with local demand conditions) as instruments to estimate child cognitive ability production functions.

Because welfare rules are complex, we need many variables to characterize them. Thus, we face a “many instrument problem” (i.e., 2SLS severely biased toward OLS). We deal with this problem both by using LIML, and by using factor analysis to condense the instrument set. Results from the two approaches are very similar, and quite different from OLS. Using LIML along with factor analysis of the instruments leads to an efficiency gain (i.e., smaller standard errors) relative to using LIML alone.

In our baseline specification, we estimate that a year of childcare reduces child test scores by 2.1% (.114 standard deviations). This estimate is quite robust across a wide range of specifications and instrument sets. But we find important interactions with type of care, maternal education and child gender. Indeed, only informal care leads to significant reductions in cognitive outcomes. Formal center-based care does not have any adverse effect. In addition, the value of the maternal time input is greater for more educated mothers, and girls are more adversely affected by childcare than boys. We do not find differential effects by child age or race/ethnicity.

¹ We acknowledge support from Australian Research Council grants FF0561843 and DP0774247, and Bernal acknowledges financial support from “The Searle Fund.” Corresponding author: michael.keane@uts.edu.au.

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 16.² 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 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.

² See, e.g., Keane and Wolpin (1997, 2001), Cameron and Heckman (1998), Cunha and Heckman (2006).

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 (2006), Andrews and Stock (2006). We deal with this problem in two ways: First, as Hansen et al note, the LIML estimator corrects the 2SLS bias in the many instrument case, so we use LIML. Our 2SLS estimates fall in between LIML and OLS, suggesting 2SLS does suffer from severe bias. Stock-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 ability test scores by roughly 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 outcomes. 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, have used the NLSY to assess effects of maternal employment and childcare use on child cognitive development. Recent reviews of this literature include Love et al (1996), Blau (1999a), Lamb (1996), Haveman and Wolf (1994), Ruhm (2002) and Blau and Currie (2004). Here we provide a brief review of 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 effects that are either insignificant or that vary by the group studied or the timing of inputs. Results in the literature 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:

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

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) skill endowment – e.g., the mother’s education, AFQT score, etc., or the child’s birthweight. The error components, μ_j and δ_{ij} are family and child effects, which capture parts of the *unobserved* skill endowment of the child. Finally, ε_{ijt} is a transitory error term that 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 actually estimate (1), and many estimate equations that seem quite far from it. One fundamental problem is that the maternal time input T and the goods inputs G are not directly observed. Most papers ignore the problem that T is unobserved, 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 μ_j and δ_{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,³ 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 μ_j but does not eliminate the child effect δ_{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.

³ 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.

Blau (1999a) and Duncan and NICHD (2003) also 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 μ_j and δ_{ij} .⁴ 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.

To our knowledge, only two prior papers attempt to use 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 possibly 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.⁵ The main advantage of our approach is that the instruments we employ are much stronger, as we'll see in Section 6.

Bernal (2006) takes a different approach by estimating a structural model of work and child care choices of *married* women. She estimates a child cognitive ability production function – which includes mother's work and 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 several reasons: First, to see if results generalize. Second, single mothers are of special policy relevance, as welfare reform led to large increases in their work/childcare use. Third, 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 recent papers also estimate cognitive ability production functions, but for children who are 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 et al. (2003) study 5 to 15 year olds, and Todd and Wolpin (2007) and Cunha and Heckman (2006) look at 6 to 13 year olds. Thus, none of these studies address how childcare affects child outcomes. The first

⁴ 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.

⁵ For this reason, James-Burdumy's preferred specification uses sibling differences to control for household fixed effects, and does not use IV.

three papers adopt conventional empirical approaches (sibling fixed effects, a structural model and a value added model, respectively). Cunha and Heckman (2006) adopt the novel approach of treating investment as a latent variable. The items of the NLSY HOME environment index are used as noisy measures of investment in children. 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.

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

Several studies have examined the relation between test scores in childhood and subsequent outcomes like educational attainment and wages. While causality is difficult to ascertain, this research shows that child cognitive achievements are strong *predictors* of a variety of outcomes in later life. This highlights the need to better understand the determinants of early test scores.

The studies that link test scores at the earliest ages to later outcomes use the British National Child Development Study (NCDS). Hutchinson et al (1979) use the NCDS to link test scores at age 7 with scores at age 16. Connolly et al (1992) find a significant positive relationship between scores at 7 and earnings at 23 (in a sample of men who left school at 16). Robertson and Symons (1996) and Harmon and Walker (1998) find a positive association between scores at age 7 and earnings at 33. And Currie and Thomas (2001) find math scores at 7 are strong predictors of math scores at 16.

All these studies look at tests taken at age 7 or older.⁶ Do tests at even earlier ages predict subsequent achievement? In Appendix 1 we present evidence from the NLSY that the PPVT at age 4 and PIAT reading and math scores at ages 5 to 6 are significantly correlated with educational attainment of youth who are at least 18 years old. For example, consider a one-point increase in the math score at age 6 (i.e., roughly a 1% increase, as the mean is 99.7). Holding background variables like mother's education fixed, this is associated with increased educational attainment (measured at age 18 or later) of approximately .019 years. Similarly, a one-point (or roughly 1%) increase in the reading score at age 6 is associated with an increase in highest grade completed of approximately .025 years. These estimated impacts are fairly substantial, since our estimates imply that a year of full-time maternal work combined with informal childcare use reduces test scores by roughly 2.6%. This translates into an effect on completed schooling of roughly .050 to .065 years, a large effect.⁷ Also, a striking aspect of the results is that mother's AFQT is not a significant predictor of

⁶ The studies we are aware of that use U.S. data (i.e., Neal and Johnson (1996), Murnane et al (1995) and Zax and Rees (1998)) look at test scores measured at age 14 or later (See Web Appendix A for details).

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

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 (2007) 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 were *entitled* to receive benefits if they met income and asset eligibility requirements. But under Section 1115 Waivers and TANF, States can set time limits on benefits. Indeed, PWRORA forbids States from using federal funds to provide benefits to adults beyond a 60-month lifetime time limit, and it allows States to set shorter limits. For instance, California sets a 5-year limit, while Texas and Florida set limits in the 2-3 year range.

Time limits have direct and indirect effects. The direct effect is simply to hit the time limit and become ineligible. The indirect effect arises if women are forward-looking and “bank” months of eligibility for later use. We use eight variables to capture both effects of time limits (see Table 1). They measure time limits created under both TANF and AFDC waivers. We include, for example, a dummy for whether a single mother's State of residence had imposed a time limit (TLL_{st}) by time t , a dummy indicating if the time limit could be binding for a particular woman (TL_HIT_{ist}), and her maximum potential remaining time of eligibility ($REMAIN_TL_ELIG_{ist}$).

It is worth emphasizing that when possible we construct instruments that are person specific. For example, consider TL_HIT_{ist} . Say a woman resides in a State that had 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 have participated in AFDC/TANF for 5 years, and therefore could not have hit the limit. Crucially, however, we do not use a woman's actual welfare participation history to determine if she had hit a time limit, because the actual history is endogenous. Our individual specific instruments are functions of policy rules and demographics alone. Regarding demographics, we assume ages of a woman and her children are exogenous (conditional on age controls in the main equation). However, as we discuss later, we treat fertility (number of children) as endogenous.

3.2. Work Requirement Time Limits and Work Requirement Exemptions

Work requirements increase time and utility costs of receiving welfare. Under PRWORA, recipients must commence “work activities” within two years to continue receiving TANF benefits. But many States adopted shorter time limits. Due to variation in when States implemented TANF, and in the length of their time limits, there is substantial cross-State variation in how early single mothers could have been subject to binding work requirements. Also, States may exempt from work single parents with children up to 1 year old. Thus, within a State, there is variation across women in whether work requirements can be binding based on age of the youngest child.

We constructed a total of nine variables, listed in Table 1, to capture these various effects. For example, WR_HIT_{ist} , is an indicator for whether the woman could have been subject to work requirements (based on the time limit, time since it had been implemented, ages of children, etc.), and $CHILD_EXEM_{st}$ is a dummy for whether state s had an age of youngest child exemption at t .

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

AFDC/TANF benefits are, roughly speaking, determined by a formula where a State specific grant level, increasing in number of children under 18, is reduced by some percent of the recipient’s income. Two variables we use to characterize the system are the maximum *potential* real monthly AFDC/TANF benefit for families with one and two children ($BEN(1)_{st}$ and $BEN(2)_{st}$), assuming zero earnings, in the mother’s State of residence. We do not condition on actual family size, as we treat fertility as endogenous.⁸ We also use the “benefit reduction rate” (BRR), the rate that benefits are reduced with income. This changed several times over the history of the AFDC program. Under waivers and TANF, the BRR was made State specific, and it now varies considerably across States.

AFDC also incorporated “earnings disregards” to encourage work – i.e., if a recipient started work, then for a period of time, a fraction of her earnings was not subject to the BRR. Generally, the disregard consisted of a “flat” part (e.g., the first \$30 of monthly earnings) and a “percentage” part (e.g., one-third of earnings beyond the flat part). Both would be eliminated after a certain number of months. Starting in late ‘92, many States obtained waivers to increase disregards. Under PRWORA, States are not required to adopt any particular disregards, so a great deal of State heterogeneity has emerged. A few States expanded disregards and allowed them to apply indefinitely. We code the BRR and the percentage disregard together in the variable $PERC_DISREGARD_{st}$. Flat disregards are coded in $FLAT_DISREGARD_{st}$.

⁸ Benefits are put in real terms using a region-specific CPI. Since 1980, the BLS computed the CPI for 24 metropolitan areas. Potential benefits in other areas were deflated using a region-specific (western, south, midwest and northeast) CPI.

3.4. Child Support Enforcement

Child support is an important source of income for single mothers, despite widespread non-payment.⁹ The Child Support Enforcement (CSE) program, enacted in 1975, implemented programs to locate absent parents and establish paternity. CSE expenditures increased significantly from \$2.9 billion in 1996 to \$5.1 billion in 2002. These expenditures are an important indication of how likely a single woman is to collect child support. We obtain a measure of State level CSE activity by dividing State CSE expenditures by the State population of single mothers ($ENFORCE_{st}$).

3.5. Child Care Subsidies and the Child Care Development Fund (CCDF)

CCDF is a block grant to States to provide subsidized childcare for low-income families. Under CCDF, states have autonomy to design their own programs, so a great deal of across-State 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.

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

The EITC is a refundable Federal tax credit that supplements wages for low-income working families. EITC was originally a minor program, but a major expansion in 1994-96 made it a sizable wage subsidy. Thus, EITC may provide an important work incentive.¹⁰ The EITC subsidy rate varies by family size. We use as instruments the EITC subsidy rates for families with one and 2+ children ($EITC(1)_{st}$ and $EITC(2)_{st}$ respectively), using Federal and State EITC rules. As with benefit levels (see Section 3.3) we do not condition on actual family size, which we view as endogenous.

Finally, we use two variables that measure 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

Leibowitz (1974) first used the human capital production framework (see Ben-Porath (1967)) to examine investments in children. Adopting this framework to analyze the development of child cognitive ability, we start by specifying a child cognitive ability production function. Letting A_{it} be child i 's cognitive ability t periods after birth, we can write the production function:

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

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 ω_i is the child's ability endowment. Goods inputs may

⁹ In 2002, child support accounted for approximately 6.5% of single mother's real incomes (March CPS).

¹⁰ For example, in 2003, the phase-in and phase-out rates for a family with *one child* were 34% and 15.98%, respectively. As of 2003, 17 States had enacted State earned income tax credits that supplement the federal credit.

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). First, estimation of a completely general specification, where inputs may have a different effect at each age t , and where the endowment ω_i may differentially affect ability at each age, is infeasible due to proliferation of parameters.¹¹ Thus, we obviously need to restrict how inputs enter (2). One simplification, familiar from the human capital literature, is to assume that: (i) only cumulative inputs matter, and (ii) the effect of the permanent unobservable is constant over time. Letting $\widehat{X}_{it} = \sum_{\tau=1,t} X_{i\tau}$ be the cumulative amount of input X up through time t , and assuming that cumulative inputs affect $\ln A_{it}$ linearly, we obtain:¹²

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

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

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

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

We now consider problems of estimating the production function in the special case of (3)-(4). The first 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 (i.e., time with the mother and time in childcare) and assume that $T_{it} = T - C_{it}$, where T is total time in a period. Then, we can rewrite (3) as:

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

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

The second key 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

¹¹ 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 $20 \cdot 21/2 = 210$ parameters for that input alone.

¹² It is convenient to let the cumulative goods input enter in log form, for reasons that will become apparent below.

cumulative income since childbirth, mother/child characteristics Z_i (which affect permanent income, as well as preferences), the child's unobserved ability $\hat{\omega}_i$,¹³ and child age. Specifically, we write:

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

where the stochastic term ε_i^g captures the mother's tastes for investment in the form of goods.¹⁴ Note that the prices of goods and childcare are not included in (6) as we cannot measure them, so in effect we assume they are fixed. Notably, one reason the prior literature failed to find good instruments for childcare is lack of sufficient price variation in the data.¹⁵ Substituting (6) and (4) into (5) we obtain:

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

Equation (7) is estimable, as all independent variables are observable. 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 ε_i^g . This appears likely. Decisions to work and use childcare may well be correlated with the child's skill endowment, and the work decision affects income as well. Furthermore, tastes for childcare may well be correlated with tastes for goods investment in children.¹⁶ 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 ε_i^g . This seems like a plausible exogeneity assumption.¹⁷ 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.

¹³ Note that the child's ability endowment may matter for two reasons: Either mothers may choose good inputs based on the child's ability (e.g., they may buy educational toys to compensate a child who is having certain learning problems) or because child ability affects the types of inputs a child demands (e.g., a high ability child may request more books).

¹⁴ This would arise due to heterogeneous preferences for child quality. ε_i^g may also be influenced by the child's tastes.

¹⁵ Equation (6) is consistent with several alternative models of investment. For example, if $\gamma_3=0$ there is a fixed level of investment determined by permanent characteristics, and the cumulative goods input grows at a rate given by γ_4 . At the other extreme, if $\gamma_1=\gamma_2=\gamma_4=0$ and $\gamma_3=1$ then demand for goods is simply proportional to current income ($G_{it}=\exp(\gamma_0) \cdot I_{it}$).

¹⁶ 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.

¹⁷ One might be concerned that States that adopted stricter welfare reform tended to have relatively high/low test scores initially. But in Web Appendix F we report means of child test scores prior to 1990 by State, broken down by whether the State subsequently implemented welfare waivers (i.e., moved towards reform early), or implemented strict or lenient welfare rules after 1996. There is no significant difference in pre-reform test scores between "strict" and "lenient" States.

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.

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:

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

where d_{1t} and d_{2t} are cognitive ability test dummies, which capture the mean differences in scores across the three tests,¹⁸ and ε_{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 appropriate, 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:

$$(9) \quad \ln S_{it} = \varphi_0 + \varphi_1 \cdot t + \varphi_2 \widehat{C}_{it} + \varphi_3 \ln \widehat{I}_{it} + \varphi_4 Z_i + \eta_1 d_{i1t} + \eta_2 d_{i2t} + \eta_3 d_{it} X_i + \nu_{it}$$

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

In our empirical work we consider many generalizations and alternative formulations of (9). For instance, we compare cumulative and current input specifications. We also 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). This captures the notion that the effect of home inputs on child’s cognitive ability might vary depending on the type of mother. And we test for differences in the effect of childcare depending on characteristics of the childcare provider (i.e., formal vs. informal) and of the child (i.e., age, race, gender).

¹⁸ 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.

5. Data

5.1. Construction of the Sample

We use data from the National Longitudinal Survey of Youth 1979 cohort (NLSY79). This contains 12,686 individuals (roughly half women), aged 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 (see Sect. 5.2). We use single mothers in the NLSY79 to estimate the child cognitive ability production function, precisely because their work/childcare decisions were greatly affected by policy changes in our sample period.

Thus, we require that women in our sample be single (or not cohabitating with a male) during five years following the birth of the 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 (an average of 2.59 per child). We use the geocode data to identify State of residence of each mother so as to construct State specific welfare rule parameters.

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 by welfare reform by age 5. 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, we also get leverage for identification by comparing children in States with “strict” vs. “lenient” welfare rules.

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 a detailed employment history for each mother in the sample for the period surrounding the birth of each child, up to four quarters before birth and 20 quarters after birth (for a period of five years). For childcare, retrospective data were gathered during 1986, 1988, 1992, and 1994-2000 that allow us to construct complete quarterly child care 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 during the last month. This is inadequate to determine if care was full or only part-time. However, by combining the childcare variable with data on work, we can make a reasonable determination about full vs. part-time care. In making this imputation, we benefit from the fact that our sample is single mothers, as we know their childcare hours must be at least as great as their work hours.¹⁹ Thus we use the following procedure:

If a woman reports using 10+ hours per week of 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 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. Thus, we define the function:

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

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

As we noted earlier, complete childcare histories are only available for the first three years after childbirth. Thus, we impute childcare choices in years 4 and 5 after childbirth based on current work and work/childcare histories. We are aided in making this imputation by the fact that childcare use and work are so closely linked for single mothers, and by the fact that there is great persistence in their childcare choices. To begin, 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. Then, for mothers who do not work in a given period t , we impute the childcare choice based on the predicted probability of using childcare from a probit model we estimate using observed work and childcare histories. As the probit coefficients in Appendix 2 show, childcare use by non-working mothers is very well predicted by lagged childcare

¹⁹ 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 do use 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. 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%.

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 or dividends, pension, 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.

An issue we did not discuss in deriving (9) is that mothers may have multiple children, which may affect resources allocated to any one child. Thus, we include the number of children in Z_i in (9). We will also try interacting childcare and income with number of children (to allow input effects to differ by number of children). Of course, number of children may be endogenous in (9), e.g., if there is a quality/quantity tradeoff, so we instrument for these variables using the instruments in Table 1.

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. The PPVT is a vocabulary test for standard American English and provides a quick estimate of verbal ability and scholastic aptitude. The PIAT-M consists of eighty-four multiple-choice items that measure math skills ranging from simple numeral recognition and to more advanced concepts in geometry and trigonometry. The PIAT-R measures word recognition and pronunciation.

5.3. Descriptive Statistics

In Table 4 we present means and standard errors of the variables used in the analysis. For example, the average log test score in the sample is 4.50 with a standard deviation of 0.186 (after adjusting for mean differences across tests). 64% of women in the sample worked prior to giving birth at an average hourly rate of \$9.14 in 2007 dollars. Average work experience was 4.7 years prior to childbirth, and 72% of women had never been married. Average annual real household income is \$22.7 thousand (2007) dollars. During the 20 quarters after childbirth mothers use .355 units of childcare per quarter, for a total of 7.1 quarters, on average. However, if we compare 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.

Appendix 3 contains descriptive statistics for test scores in our sample. Note that there is no clear age pattern in the mean scores, as they are age adjusted. Mean scores on the PPVT, PIAT-M and PIAT-R are roughly 80, 95 and 101.²⁰ Interestingly, score differentials between children who are white/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 correlation²¹ squared. 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 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.²² 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

²⁰ 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.

²¹ This partials out the correlation of an endogenous variable with fitted values of the other endogenous variables.

²² These contain 101 variables, of which 23 are exogenous variables also appearing in the main equation, and 78 are excluded instruments. As the main equation contains 3 endogenous variables, there are 75 over-identifying restrictions.

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 significance 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.

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.

Next, we assess whether a cumulative or current childcare specification is most appropriate. Web Appendix D reports estimates of both models. The LIML estimate of the effect of a quarter of childcare on test scores is -2.7% . The cumulative specification implies the effect of a quarter of childcare is only one-fifth as great. Given the persistence in childcare choices noted earlier, it is not surprising for current childcare to proxy for lagged childcare. If we add four years of lagged childcare indicators to the current specification, the p-value for their joint significance is $.0329$. Thus, we reject that only current childcare matters. Also, a χ^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. Here we report estimates based on several alternative estimation methods. The OLS estimate of the effect of cumulative childcare is essentially zero and insignificant. In contrast, 2SLS based on the 78 welfare policy and local demand condition instruments produces an estimate of -0.36% per quarter (or -1.4% per year) that is significant at the 10% level. But, as noted earlier, 2SLS based on such a large number of excluded instruments is likely to be severely biased towards OLS. Turning to LIML, we obtain an estimate of -0.52% per quarter (or -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 almost no effect on the LIML point estimate, but it reduced the standard error sufficiently that it is 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 Appendix 1, a 2.1% test score reduction at age 6 translates into a .040 to .053 year reduction in completed schooling.²³

Table 6 also reports estimates for cumulative income, mother's education and AFQT. The estimated effect of income since childbirth is quantitatively small and statistically insignificant. For instance, the point estimate of .0106 in the last column of Table 6 implies that, at the mean of the data, *doubling* cumulative income (i.e., increasing its log by .69 – see Table 4) would increase test scores by only $(.0106) \cdot (.69)=0.7\%$. In contrast, mother's education and AFQT are highly significant and quantitatively important. This is *consistent* with a view that lifetime income is much more important than transitory income in determining parental investment in children, and hence child achievement.²⁴ But of course mother's education and AFQT may also be important for other reasons: genetic transmission of ability, more educated mothers may be better at the technology of teaching children, etc.. We do not attempt to disentangle these alternative mechanisms.

6.3. Comparison of Alternative Estimation Methods

As we noted in the introduction, 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 is shifted about 25% of the way towards OLS. This is not surprising, as with 78 instruments, we would expect 2SLS to retain some part of the OLS endogeneity bias.

The bottom row of Table 6 reports the Cragg-Donald (1993) weak instrument test statistic, which is 5.80 when the full set of 78 instruments is used. Stock and Yogo (2004) develop critical values of this statistic, for testing the null hypothesis that the asymptotic maximal bias of the 2SLS estimator may exceed some percentage of the OLS bias (under many instrument asymptotics). For the case of 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.

²³ Repeating the calculation at the end of Section 2, this gives about a 6% decrease in the number who attend college.

²⁴ 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 in determining investments in children while transitory income is not.

Thus, we cannot reject the null that the 2SLS asymptotic bias may exceed 10% of the OLS bias, and we only barely reject the null that it may exceed 20% (i.e., 5.80 vs. 5.65).

Next consider the LIML and Fuller estimators, which produce almost identical results. While Stock and Yogo (2004) do not report exact values for all cases, we extrapolate from their figures that the critical value for the null that bias of the Fuller estimator is no greater than 5% of the OLS bias is roughly 1.8 in the 78 instrument case – easily exceeded by our 5.80 value for the Cragg-Donald statistic. And the critical value for the null that bias in size for the LIML test statistics is no greater than 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 important efficiency losses (i.e., standard errors one-third greater than 2SLS). Thus, we tried an alternative approach to the many instrument problem based on condensing the size of the instrument set. Specifically, we summarize the information contained in the instruments using a smaller number of variables obtained via factor analysis. To do this we used the principal factor method with varimax rotation. The factor scoring coefficients are calculated via the regression method. It is important to emphasize that the estimated factors constructed using this method are simply linear functions of the original 78 instruments. Thus, assuming the original instruments are valid, the estimated factors will be valid as well.

A typical 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, were 6, 8, 12, 19, 21, 24 and 26. Given which variables load on each factor, we interpret them as follows: Factor 6 primarily captures benefit levels. Factor 8 captures remaining eligibility, local wages and EITC. 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.²⁵ Factor 24 captures CCDF spending, CSE and EITC. Factor 26 captures benefit levels, remaining eligibility, EITC and local wages.

²⁵ 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.

For cumulative income, the most important factors were 1, 3, 6, 7, 10, 24 and 26, while for number of children, they were 1, 2, 3, 6, 8, 9, 12, 21, 23 and 26. (In the interest of space we will not provide interpretations of all these factors). There is overlap among the important factors for each of the three endogenous variables. Taking the union of the sets of factors that matter for each variable, we obtain 14 variables to use as instruments, thus reducing the instrument set from 78 to 14.

Table 7B examines correlation of the 14 factor instruments with cumulative childcare. For instance, Shea's partial correlation squared is .0967 (vs. .1483 for the full set of 78 instruments), and the F-test of their joint significance is 14.57 ($p=.0000$). The factors have sensible coefficients in the reduced form equation for cumulative childcare use (not reported). 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 use. And factor 21 has a positive coefficient ($t=7.6$), suggesting that work experience/cumulative childcare grows more quickly with child age for mothers with more prior experience. [Notice that the factors with the most explanatory power for childcare do not correspond at all with those that had the largest eigenvalues in the factor analysis].

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 14 factors in place of the full set of 78 instruments. Second, and most importantly, using the factors as instruments does lead 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, it is notable that the LIML and 2SLS estimates are very similar once 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 (compared to 25% when we use all 78 instruments). Indeed, the Cragg-Donald weak instrument test statistic, which is 5.80 when the full set of 78 instruments is used, increases to 15.33 when we use the reduced set of 14 instruments (see the bottom row of Table 7A). 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 clearly reject the null that the 2SLS asymptotic bias may exceed 10% of the OLS bias, but we do not quite reject the null that it may exceed 5%. These results suggest that using factor analysis to reduce the size of the instrument set is a rather effective way to reduce the 2SLS bias.

The last column of Table 7A reports results using the first 21 principle factors as instruments, instead of choosing factors that best predict the endogenous variables. This gives a slightly larger estimate of the childcare effect (-0.63% per quarter, $t=-2.3$). However, four disadvantages of this approach are apparent: First, as we show in Web Appendix E, it requires more factors (21 vs. 14) before results “settle down.” Second, it achieves little efficiency gain – the standard error is little different from LIML based on all 78 instruments. Third, the Cragg-Donald statistic drops to 11.13 vs. 15.33 in our procedure. Fourth, as we see in Table 7B, these 21 factors have less explanatory power for the endogenous variables than our 14 factors. Given a large set of valid instruments, it seems natural to choose, as we do, those most highly correlated with the endogenous variables.

6.4. Robustness of the Results with Respect to Age of the Mother at Childbirth

A potential concern is 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. Say a woman had a child in 1990, reaching age 5 in 1995. Her work/childcare decisions could have been impacted by waivers when the child was age 5, possibly affecting test scores at 5 and 6. But, for a child born prior to 1990, it is unlikely waivers could have influenced the mother’s labor supply before he/she 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.²⁶

To understand the potential bias created by this correlation, consider the following scenario: Since work requirements positively affect maternal work/childcare use, and work requirements are positively correlated with age at childbirth, our reduced form predicted values for maternal work/childcare use will be *positively* correlated with maternal age at childbirth. Then, if (i) mother’s age at birth has a positive effect on child cognitive ability, and (ii) we fail to adequately control for mother’s age in the main equation, this will generate a spurious *positive* effect of maternal work/childcare use on child cognitive test scores. Thus, as we actually find a negative effect of maternal work/childcare use, the concern is that we *understate* this negative effect.

To deal with this concern, Table 8 presents several specifications of the main equation. The results in columns (1)-(5) show the estimated effect of cumulative childcare use is very robust to the inclusion/exclusion of different controls for mother’s age at childbirth – including age, age squared, and dummies for if she was under 20 or over 33. In fact, the estimated effect ranges from -0.49% per

²⁶ Indeed, the youngest women in the NLSY – i.e., those who were 14 in Jan. 1979 – would be 24 by Jan. 1990. Thus, the large majority of mothers who would have been affected by welfare reform would have been at least 24 at childbirth.

quarter with no controls for age at all, to -0.52% with all four controls. The estimate is -0.53% in the baseline model in column (4), which includes only the teenager and over 33 indicators.

A striking aspect of the results is that, conditional on measures of the mother's human capital (i.e., education, AFQT), there is no evidence of a *positive* association between maternal age at birth and child cognitive outcomes. Indeed, the column (4) estimates imply that, *ceteris paribus*, children of teenage mothers have 2.4% *higher* test scores. It would appear that, controlling for economic resources, maternal youth is beneficial for child outcomes.

To further address this issue, in Table 9 we also report estimates for sub-samples of women based on age at childbirth. We restrict the sample to women who were 24+, 24-34 or 24-30 years old at childbirth. For these sub-samples, the estimated effects of childcare range from -2.3% to -1.6% per year. However, due to the reduced sample size these estimates are not significant. So in columns (4)-(6) we use the full sample, but we interact the childcare coefficient with dummies for whether the mother's age at childbirth was outside the indicated range. None of the interactions is significant, and the main effect of childcare is quite stable across the three columns (i.e., it ranges from -2.3% to -2.4% per year and is always significant). Thus, we find no evidence that the age/welfare policy correlation leads to significant bias in our estimates of the childcare effect.

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

We now return to Table 8 and consider sensitivity of our results to five other changes in the specification of the main equation. First, we drop the mother's AFQT score. This leads to a slight increase in the estimated childcare coefficient (to -2.4% vs. -2.1% per year in the baseline). More noticeable is that it produces a large increase in the cumulative income coefficient, which becomes highly significant. This seems 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.

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 would increase test scores by about $(.078) \cdot (.69) = 5.4\%$. Recall that the model still includes such variables as mother's education and her pre-childbirth wage, which also proxy for her permanent income/skill endowment.

Next, we examine if maternal work has a separate effect from childcare. *Married* women often use childcare when not working, so one can estimate effects of work time holding childcare time fixed (see Bernal, 2006). But use of childcare while not working is much less common for

single mothers. The correlation between cumulative work and childcare is very high ($\rho=.94$), making it nearly impossible to separate their effects. Instead, we construct a measure of work equal to 1 if a woman works continuously after childbirth, 0.5 if she works part of the time, and 0 if she did not work at all. This variable (which we treat as endogenous) has a coefficient of -0.073, and with its inclusion, the childcare effect falls to -0.32% per quarter. But both estimates are very imprecise. Clearly, collinearity makes it infeasible to sort out effects of work vs. childcare for single mothers.

Third, we consider sensitivity of our results to controls for the ages of siblings. Our baseline model controls for the number of siblings, not their ages. If younger children have a different effect on the mother's time constraint, this may bias our estimated childcare effect. Thus, the next column includes separate regressors 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 (9) 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.

Finally, in the last column of Table 8 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.

6.6. Robustness of the Results with Respect to State Fixed Effects

The argument for including State effects is to deal with potential cross-State correlation between the instruments and *unobserved* child skill endowments; e.g., States where children had relatively low unobserved skill endowments may have adopted stricter welfare reform. This would bias our estimated childcare effect negatively. To deal with this concern we report State fixed effects estimates in Table 10. Adding State fixed effects to the main equation actually shifts our estimate of the cumulative childcare effect from -.53% to a rather implausibly large negative value of -.98%. It also reduces the precision of the estimate, more than doubling the standard error. Despite this, the estimate remains significant at the 10% level ($t = -1.85$). These results suggest that failure to account for state fixed effects is not biasing out estimated childcare coefficient in a negative direction.

But we hasten to add that we are skeptical of fixed effects for two reasons: First, in the child production function context, the strict exogeneity assumption required for consistency is not plausible.²⁷ Second, Keane and Wolpin (2002) show fixed effects can lead to very misleading results if expected future values of policy variables matter for current decisions. In our view the implausibly large fixed effects estimate probably reflects the difficulty in knowing what fixed effects actually estimates in a context where (i) strict exogeneity is implausible and (ii) changes in childcare driven by *transitory* policy changes may have a very different impact from changes driven by *persistent* changes.²⁸ Furthermore, as we see in the bottom panel of Table 10, State effects are not significant at the 5% level. Given these considerations, we do not adopt fixed effects as the baseline model.

An alternative way to see if States where children had low skill endowments adopted stricter welfare reform is to look at the issue directly. In Web Appendix F we group States into those that adopted more vs. less strict approaches to welfare reform along five different dimensions. Then we compare average test scores in the pre-reform period between each group of States. In each case, there is no significant difference in average child test scores in the pre-reform period between States that subsequently adopted more vs. less strict welfare reform programs. Thus, there is no clear evidence of cross-State correlation between the instruments and child skill endowments.

6.7. 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 11 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. This shifts the effective price of childcare, so it 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 policy and demand environment to identify the childcare effect. This reduces the estimate slightly, to -1.7% per year.

²⁷ The strict exogeneity assumption will fail if children's test score realizations at age t affect future inputs into child production, and/or how the welfare policy rules evolve. [See also the discussion of this point in Section 2].

²⁸ A State fixed effect controls for a State's average level of welfare generosity. Thus, using State effects, we estimate the impact of a transitory change in work/childcare use induced by a transitory change in welfare rules, holding a State's expected future rules fixed. This effect may be very different from that induced by persistent lived policy changes.

In column (5) we drop all instruments specific to the welfare reforms of the 90s (e.g., TANF, CCDF, EITC), using only those that varied across States/time regardless. These are State welfare grant levels and local demand conditions (i.e., State unemployment and 20th percentile wages). Here the childcare effect estimate is slightly smaller than in our baseline, -1.75%, and it is significant at the 10% level ($t=-1.82$). This is our most parsimonious model, with only 18 instruments.

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 almost no impact on the estimated childcare effect.

Recall that some of our instruments are tailored to individuals based on ages of their children (e.g., whether a woman could have reached the time limit – see Section 3.1). In column (7) we drop these individual specific instruments, and rely purely on cross-State and over time variation to identify the childcare effect. 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 the interactions of the remaining instruments with mother's education and AFQT. This gives an estimate of -2.6% per year.

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 have experimented with a large number of other instrument sets (not reported) and continue to find results in this general range.

6.8. Heterogeneity in the Effect of Maternal Time Inputs

In Table 12 we assess how the effect of childcare varies with characteristics of the mother or household. In the first three columns we include interactions between cumulative childcare and mother's education, AFQT and number of children. These variables are de-measured before being interacted with cumulative childcare. Thus, the estimates of the main effects can be interpreted as the affect of cumulative childcare for a typical mother/household.

In column (1) note that the *interaction* between mother's education and childcare use is *negative*, as expected (maternal time is a less valuable input into child cognitive ability production for less educated mothers). Its t-statistic is -1.76, so it only attains significance 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 -2.90.²⁹ Thus, we have strong evidence that

²⁹ The reason this is so highly significant is due to the negative covariance between the main effect and interaction term.

childcare has more negative (*positive*) effects for children whose mothers have higher (lower) levels of education. As we see in column (2), the same pattern holds for AFQT (i.e., maternal time is more valuable for high AFQT mothers). However, in column (3), we see that the interaction between cumulative childcare and number of children in the household is very small and insignificant, implying the childcare effect does not depend on number of siblings.

Column (4) includes an interaction of childcare with gender (male=1). The interaction 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). The interaction is not significant, although the point estimates imply a larger negative effect for whites. A similar result is obtained in column (6) where the non-white dummy is broken down into separate black and Hispanic dummies. Thus, we find no evidence of differences by race. Finally, the last column allows the effect of childcare to differ by child age, by including an age-weighted childcare variable.³⁰ This variable is insignificant, so we find no evidence that childcare effects differ by child age.

6.9. 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, preschool) 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.³¹ 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 for the excluded instruments in the reduced form regressions ranges from .086 to .099, and all the joint F-tests show the instruments are highly significant.

The reduced form regression coefficients (not reported) also appear reasonable. They show mothers are more likely to use formal relative to informal care if: (i) a State does *not* have a work requirement, (ii) it has young child or other work requirement exemptions, (iii) it has a longer work

³⁰ To let the childcare effect differ by age, we replace the term $\varphi_2 \widehat{C}_t$ in eqn. (9) by the term $\sum_{\tau=1,t} (\varphi_{20} + \varphi_{21} \cdot \tau) \cdot C_t$. This lets the impact of childcare be a linear function of child age. This expression may be rewritten $\varphi_{20} \widehat{C}_t + \varphi_{21} \cdot \sum_{\tau=1,t} (\tau \cdot C_t)$, where $\sum_{\tau=1,t} (\tau \cdot C_t)$ is an “age weighted cumulative child care” variable which we now add to the regression and treat as endogenous. In Table 5 Panel B we see that the incremental R^2 for the excluded instruments in the reduced form regressions for this variable is .103, with a F-statistic of 25.3.

³¹ 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 2. 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%.

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. Interestingly, work requirements raise the probability of using informal, but not formal care. And 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 are always opposite in sign to main effects: welfare rules have less influence on more educated women. A high AFQT score reinforces the effects that young child exemptions and higher disregards increase the likelihood of using formal care. Only for high AFQT women does the probability of using formal care increase in the time since time limits were implemented.

Strikingly, the LIML results in Table 13 indicate that formal (i.e., center-based) care does not have any adverse effect on cognitive outcomes. Only informal types of care lead 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 G, and they show that the result is quite robust to changes in specification, the instruments, etc..

In column (5) of Table 13 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 prior work by Hansen and Hawkes (2007). 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 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 Lavery (2003)). Prior literature also suggests center-based care has two advantages over informal care: (1) trained care providers may provide more cognitive stimulation to children than informal providers,³² and (2) center-based care may provide more stimulating interaction with other children and more educational activity than informal care.³³

In the last column of Table 13 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., +0.8% 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 et al (2006) on the Perry Pre-School and Abecedarian experiments).

6.10. The Effect of Welfare Reform on Child Test Scores

Our focus has been on using the welfare policy changes on the mid-90s as a source of exogenous variation to help identify the effect of childcare time vs. maternal time on child outcomes. But it is also interesting to examine how welfare reform itself affected child outcomes. We do this using the reduced form equation for child test scores, obtained by substituting all the excluded instruments listed 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 these factors to welfare reform.

The 78 instruments are highly significant in the reduced form test score equation – the F-test for their joint significance is 2.79 compared to the 1% critical value of 1.44. In a simpler model that leaves out interactions with mother’s education and AFQT, the remaining 30 instruments give an F-test of 2.20, compared to the 1% critical value of 1.70. These results suggest that changes in welfare

³² 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.

³³ Ideally, we would also like to examine how effects of childcare differ by direct measures of childcare quality. But it is difficult to measure quality directly in the NLSY. Lacking a direct quality measure, 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 H we present a logit for whether a mother uses formal or informal childcare (conditional on child care use). The results show that more educated, urban women with fewer children are more likely to use formal care. This suggests that formal care is higher quality, as it is typically used by women who can afford more expensive care. Similarly, Web Appendix H 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.

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. Thus, we instead simulate the effect of changes of welfare rules by simulating test scores from the reduced form. We simulate 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. A similar procedure was used by Fang and Keane (2004) to evaluate effects of changes in welfare rules on employment and welfare participation by single mothers.

Simulation of the reduced form model implies that changes in welfare rules had almost no impact on child test scores during the 1979-93 period. This is not surprising, as the rule changes during that period were modest. However, the rule changes began to reduce test scores after 1993. Our model implies that average test scores for children of single mothers in the 1994-99 period were 1.32% lower than they would have been had the rules not changed.

This figure is broadly consistent with our point estimate for the effect of child care. As we see in Table 4, the child care usage rate was about 10 percentage points higher in the post-93 period. Thus, by 1999 (6 years later) children would have had about 0.6 extra years of child care on average. Multiplying this by our estimated annual effect of -2.13%, we obtain -1.3%, which is quite close to our simulated effect of -1.32%.

6.11. Childcare and Non-Cognitive Outcomes

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

The OLS point estimate implies essentially no effect of childcare on the incidence of behavioral problems. However, the LIML results suggest a quantitatively large adverse effect of +0.46% per quarter of childcare.³⁴ But the t-statistic on this estimate is only 1.44, the low level of significance presumably 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.

³⁴ 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 outcomes 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 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.

References

- Anderson, T.W. and H. Rubin, 1949, "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46-63.
- Anderson, T.W., N. Kunitomo and Y. Matsushita, 2005, "New Light from Old Wisdoms: Alternative Estimation Methods of Simultaneous Equations in Microeconomic Models," CIRJE Discussion paper #F-321, University of Tokyo.
- Andrews, D.W. and J. Stock, 2006, "Inference with Weak Instruments," in Blundell R., W. Newey and T. Persson (eds.), *Advances in Economics and Econometrics*, Vol. 3, Ch. 6, Cambridge University Press.
- Ben-Porath, Y., 1967, "The production of human capital and the life-cycle of earnings," *Journal of Political Economy*, vol. 75, no. 4, part I, August, 352-265.
- Bernal, R., 2006, "The Effect of Maternal Employment and Child Care on Children's Cognitive Development," *International Economic Review*, forthcoming.
- Bernal, R. and M. P. Keane, 2007, "Quasi-Structural Estimation of a Model of Child Care Choices and Child Cognitive Ability Production," manuscript, University of the Andes.
- Blau, D., 1999a, "The Effects of Child Care Characteristics on Child Development," *The Journal of Human Resources*, XXXIV, 4.
- Blau, D., 1999b, "The Effect of Income on Child Development," *Review of Economics and Statistics*, 81(2): 261-276, May.
- Blau, F. and A. Grossberg, 1992, "Maternal Labor Supply and Children's Cognitive Development," *The Review of Economics and Statistics* 74(3), 474-481.
- Blau, D. and J. Currie, 2004, "Preschool, Day Care, and Afterschool Care: Who's Minding the Kids" in *Handbook of the Economics of Education*, Hanushek, E. and F. Welch eds. Elsevier.
- Cameron, S. and J. J. Heckman, 1998, "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy*, University of Chicago Press, vol. 106(2), pages 262-333, April.
- Carneiro, P. and J. Heckman, 2002, "The Evidence on Credit Constraints in Post-Secondary Schooling," *Economic Journal*, 112: 705-34.
- Chase-Lansdale, P.L., R. Moffitt, B. Lohman, A. Cherlin, R. Coley, L. Pittman, J. Roff, E. Votruba, 2003, "Mothers' Transitions from Welfare to Work and the Well-Being of Preschoolers and Adolescents," *Science*, 299(March):1548-1552.
- Connolly, S., J. Micklewright and S. Nickell, 1992, "The Occupational Success of Young Men who Left School at Sixteen," *Oxford Economics Papers*, 44, pgs. 460-479.
- Cragg, J.G. and S.G. Donald, 1993, "Testing Identifiability and Specification in Instrumental Variables Models," *Econometric Theory*, 9, 222-240.
- Cunha, F. and J. Heckman, 2006, "Formulating and Estimating the Technology of Cognitive and Noncognitive Skill Formation," Working paper.

- Currie, J. and D. Thomas, 1995, "Does Head Start Make a Difference?," *American Economic Review*, 85:3, 341-364.
- Currie, J. and D. Thomas, 2001, "Early Test Scores, School Quality and SES: Long Run Effects on Wage and Employment Outcomes," *Worker Wellbeing in a Changing Labor Market*, Vol. 20.
- Duncan, G. and NICHD Early Child Care Research Network, 2003, "Modeling the Impacts of Child Care Quality on Children's Preschool Cognitive Development," manuscript Northwestern University.
- Fang, H. and M. Keane, 2004, "Assessing the Impact of Welfare Reform on Single Mothers," *Brookings Papers on Economic Activity* Vol. 1.
- Goodfellow, J. and J. Lavery, 2003, *Grandcaring: Insights into grandparents' experiences as regular child care providers*. Canberra: Early Childhood Australia.
- Gregg P., Washbrook E., Propper C., Burgess S., 2005, "The effects of a mother's return to work decision on child development in the UK," *The Economic Journal*, 115:F48-F80.
- Hansen, C., J. Hausman and W. Newey, 2006, "Estimation with Many Weak Instruments," forthcoming, *Journal of Business and Economic Statistics*.
- Hansen, Kirstine and Denise Hawkes, 2007, "The Relationship between Childcare and Cognitive Outcomes for Children of Working Mothers", Working Paper Centre for Longitudinal Studies, Institute of Education, University of London.
- Harmon, C. and I. Walker, 1998, "Selective Schooling and the Returns to the Quantity and Quality of Education," manuscript Department of Economics, University College Dublin.
- Haveman, R. and B. Wolfe, 1994, *Succeeding Generations: On the Effects of Investments in Children*. New York: Russell Sage Foundation.
- Heckman, J. and E. Vytlacil, 2005, "Structural Equations, Treatment Effects and Econometric Policy Evaluation," *Econometrica*, 73(3): 669-738.
- Hutchinson, D., H. Prosser and P. Wedge, 1979, "The Prediction of Educational Failure," *Educational Studies* 5(1).
- James-Burdumy, S., 2005, "The Effect of Maternal Labor Force Participation on Child Development", *Journal of Labor Economics* Vol. 23(1): 177-211.
- Keane, M.P. and K. Wolpin, 1997, "The Career Decisions of Young Men," *Journal of Political Economy*, University of Chicago Press, vol. 105(3), pages 473-522, June.
- Keane, M.P. and K. Wolpin, 2001, "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment," *International Economic Review*, vol. 42(4), pages 1051-1103, November.
- Keane, M. and K. Wolpin, 2002, "Estimating Welfare Effects Consistent with Forward Looking Behavior," *Journal of Human Resources*, 37:3, 570-622.
- Lamb, M., 1996, "Effects of Nonparental Child Care on Child Development: An Update," *The Canadian Journal of Psychiatry*, 41: 330-342.
- Leibowitz, A., 1974, "Home investments in children," *Journal of Political Economy*, 82:2, Part I, 111-131.

- Liu, H., T. Mroz and W. van der Klaauw, 2003, "Maternal Employment, Migration, and Child Development", manuscript University of North Carolina at Chapel Hill.
- Love, J.M., P. Schochet and A. Meckstroth, 1996, "Are They in Real Danger? What Research Does -And Does not- Tell Us About Child Care Quality and Children's Well-being", *Mathematica* No. ED145 030.
- McCartney, K., 1984, "The Effect of Quality of Day Care Environment upon Children's Language Development", *Developmental Psychology* 20:244-260.
- Melhuish, E., A. Mooney, E. Hennesy and S. Martin, 1992, "Characteristics of Child Care in Early Childhood and Child Development in Middle Childhood" presented at the European Conference on Developmental Psychology in Seville, Spain. September.
- Millward, C., 1998, *Family Relationships and Intergenerational Exchange in Later Life*, WP #15, Australian Institute of Family Studies, Melbourne.
- Moffitt, R., 1992, "Incentive Effects of the U.S. Welfare System: A Review," *Journal of Economic Literature*, 30(1), pp. 1-61.
- Murnane, R. J., J. Willet and F. Levy, 1995, "The growing importance of cognitive skills in wage determination," *Review of Economics and Statistics*, LXXVII, pp. 251-266.
- Neal, D. and W. Johnson, 1996, "The Role of Premarket Factors in Black/White Wage Differences," *Journal of Political Economy*, 104(5), pgs. 869-895.
- NICHD Early Child Care Research Network, 2000, "The Relation of Child Care to Cognitive and Language Development", *Child Development* 71(4): 958-978.
- Pungello, E., F. Campbell and W.S. Barnett, 2006, "Poverty and Early Childhood Educational Intervention," Center on Poverty, Work and Opportunity Policy brief Series.
- Robertson, D. and J. Symons, 1996, "Do Peer Groups Matter? Peer Group Versus Schooling Effects on Academic Attainment," Centre for Economic Performance, London School of Economics, WP No. 311.
- Rosenzweig, M. and P. Schultz, 1983, Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight. *The Journal of Political Economy*, Vol. 91:5, 723-746.
- Ruhm, C., 2002, "Parental Employment and Child Cognitive Development", NBER WP No. 7666.
- Stock, J. and M. Yogo, 2004, "Testing for Weak Instruments in Linear IV Regression," Working paper, Harvard University Department of Economics.
- Todd, P. and K. Wolpin, 2007, "The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps," *Journal of Human Capital*, 1(1), 91-136.
- Zax, J. and D. Rees, 1998, "Environment, Ability, Effort and Earnings", Center for Research on Economic and Social Policy, Denver, Working Paper No. 9801.

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
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)	
$BEN(1)_{st}$	Real AFDC/TANF maximum benefits for a family with 1 child
$EITC(1)_{st}$	EITC phase in rate constructed from both the federal and state level for a family with 1 child
$BEN(2)_{st}$	Real AFDC/TANF maximum benefits for a family with 2 children
$EITC(2)_{st}$	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 these policy variables and local demand conditions 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.

Table 2
Control Variables in the Cognitive Ability Production Function

Variable	Description
Baseline Specification	
$I[AGE_i < 20]$	Dummy for whether mother is younger than 20 years old
$I[AGE_i \geq 33]$	Dummy for whether mother is older than 33 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
$I[WORK_BEF]_i \times SKILL_i$	Work dummy interacted with mother's skill*
$EXPBEF_i$	Mother's total work experience (in number of years) prior to childbirth
$EXPBEF_i * age_i$	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
$NUMCHILD_{0-5}_i$	Number of children 0-5 years of age
$NUMCHILD_{6-17}_i$	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
$C_{it} * NON-WHITE_i$	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 (5.56)	23.56 (5.07)	23.80 (5.15)	23.13 (4.59)
Mother's education at childbirth (in years)	12.0 (2.475)	11.3 (1.920)	11.3 (1.917)	11.2 (1.909)
Mother's AFQT score	37.9 (27.23)	21.7 (20.09)	19.9 (19.11)	19.3 (18.30)
Hispanic or Black	0.47 (0.499)	0.73 (0.445)	0.79 (0.404)	0.83 (0.379)
Hourly wage before childbirth (first child) ^{&}	6.32 (7.71)	4.74 (8.23)	4.90 (9.85)	4.39 (2.01)
Total number of children of mother	2.9 (1.37)	3.1 (1.57)	3.1 (1.61)	3.1 (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 (0.1861)*	Urban	0.8189 (0.3851)
Mother's education	11.208 (1.8972)	Average yearly income (Thousands) ²	10.9274 (13.568)
Mother's age	23.136 (4.5820)	Cumulative income (Thousands) ³	51.1787 (67.415)
Boys (Children of single mothers)	0.4976 (0.5001)	Units of child care per quarter ⁴	0.3546 (0.3064)
Hispanic or Black	0.8262 (0.3790)	Cumulative child care use (Quarters)	7.0923 (6.1273)
Birthweight	111.97 (21.976)	Labor participation rate (avg 1979-1993)	48.23 (6.5)
Mother worked before giving birth	0.6431 (0.4792)	Labor participation rate (avg 1994-1999)	60.40 (5.4)
Wage rate prior to giving birth ¹	4.3938 (2.0075)	Welfare participation rate (avg 1979-1993)	58.93 (4.1)
Accumulated work experience prior to giving birth (number of years)	4.7202 (6.0088)	Welfare participation rate (avg 1994-1999)	44.65 (12.1)
Never married after childbirth	0.7215 (0.4483)	Childcare use rate (avg 1979-1993) ⁵	59.05 (5.0)
Separated after childbirth	0.1540 (0.3611)	Childcare use rate (avg 1994-1999)	69.27 (5.9)
Divorced after childbirth	0.1158 (0.3201)		

*Standard error of log(test score) calculated after taking out the test-specific means of the three tests, i.e., the standard error 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
<u>A. Endogenous variables in the baseline model</u>						
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	0.3252	0.3252	0.2469	0.2970	25.220	0.0000
<u>B. Other endogenous variables in additional models</u>						
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) plus mother's age and age squared, all policy variables (see Table 1), policy variables interacted with mother's education and mother's AFQT, unemployment rate in State of residence and average hourly wage at 20th percentile of wage distribution in State of residence and both interacted with mother's education and 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 (0.0008)	-0.00389 (0.0015) **	-0.00357 (0.0021) *	-0.00521 (0.0028) *	-0.00522 (0.0028) *	-0.00533 (0.0025) **
Log(Cumulative Income)	-0.00324 (0.0057)	0.01142 (0.0087)	0.00742 (0.0164)	0.01035 (0.0242)	0.01037 (0.0242)	0.01062 (0.0266)
Mother's education	0.01051 (0.0026) **	0.01130 (0.0023) **	0.01214 (0.0029) **	0.01276 (0.0032) **	0.01276 (0.0032) **	0.01297 (0.0030) **
Mother's AFQT	0.00059 (0.0002) **	0.00057 (0.0002) **	0.00063 (0.0003) **	0.00066 (0.0003) **	0.00066 (0.0003) **	0.00066 (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

Instruments used are all 78 policy variables, local demand conditions and interactions described in footnote in Table 5.

^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 asymptotic normality of LIML depend on normality, but small sample properties are presumably improved by the approximation being reasonably accurate.

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

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

[&] Fuller parameter=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 principal
	78 instruments	14 rotated factors				factors ^a
Cumulative Child Care	-0.00522 (0.0028) *	-0.00533 (0.0025) **	-0.00531 (0.0025) **	-0.00498 (0.0024) **	-0.00420 (0.0023) *	-0.00633 (0.0027) **
Log(Cumulative Income)	0.01037 (0.0242)	0.01062 (0.0266)	0.01060 (0.0265)	0.01021 (0.0243)	0.01061 (0.0213)	-0.00094 (0.0269)
Mother's education	0.01276 (0.0032) **	0.01297 (0.0030) **	0.01297 (0.0030) **	0.01282 (0.0030) **	0.01308 (0.0029) **	0.01298 (0.0033) **
Mother's AFQT	0.00066 (0.0003) **	0.00066 (0.0003) **	0.00066 (0.0003) **	0.00065 (0.0003) **	0.00063 (0.0003) **	0.00077 (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

Instruments used in first column are all 78 policy variables, local demand conditions and interactions described in footnote in Table 5.

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.[&] Fuller parameter=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 7A)

Dep var-> Cum. 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

R² 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)

	Mother's Age Controls					Removing AFQT	Include Maternal Employment	Children by ages	Year effects	Test Score in levels ^a
	None	Age	Age, Age ²	Age dummies	All age variables					
Cumulative Child Care	-0.00488 (0.0026) *	-0.00499 (0.0026) *	-0.00522 (0.0026) **	-0.00533 (0.0025) **	-0.00522 (0.0026) **	-0.00613 (0.0031) **	-0.00324 (0.0062)	-0.00553 (0.0026) **	-0.00490 (0.0026) *	-0.50258 (0.2118) **
Log(Cumulative Income)	0.00924 (0.0252)	0.01806 (0.0260)	0.00449 (0.0278)	0.01062 (0.0266)	0.00334 (0.0278)	0.07808 (0.0284) **	0.00568 (0.0325)	0.00564 (0.0280)	0.00228 (0.0300)	1.20063 (2.2509)
Mother's education	0.01133 (0.0028) **	0.01188 (0.0028) **	0.01331 (0.0031) **	0.01297 (0.0030) **	0.01340 (0.0032) **	0.01355 (0.0036) **	0.01234 (0.0036) **	0.01538 (0.0044) **	0.01329 (0.0031) **	1.05541 (0.2532) **
Mother's AFQT score	0.00071 (0.0003) **	0.00060 (0.0003) **	0.00069 (0.0003) **	0.00066 (0.0003) **	0.00069 (0.0003) **		0.00071 (0.0004) **	0.00067 (0.0003) **	0.00067 (0.0003) **	0.07814 (0.0271) **
Child's age	0.03886 (0.0119) **	0.03546 (0.0124) **	0.03977 (0.0126) **	0.03817 (0.0122) **	0.04008 (0.0126) **	0.01833 (0.0128)	0.03783 (0.0121) **	0.03812 (0.0124) **	0.04078 (0.0131) **	3.38379 (1.0073) **
Mother's age		-0.00217 (0.0016)	-0.01694 (0.0113)		-0.01346 (0.0149)	-0.00101 (0.0169)				
Mother's age squared			0.00031 (0.0002)		0.00025 (0.0003)	-0.00005 (0.0004)				
I[age of mother _i <20]				0.02368 (0.0116) **	0.00826 (0.0149)	0.00811 (0.0157)	0.02415 (0.0117) **	0.04335 (0.0272)	0.0089 (0.0115)	1.83958 (0.9577) *
I[age of mother _i >=33]				0.0060 (0.0256)	-0.0030 (0.0310)	0.0032 (0.0347)	0.0125 (0.0319)	0.0109 (0.0269)	0.0003 (0.0257)	0.2064 (2.2426)
Maternal employment ^{&}							-0.0727 (0.2081)			
Number of children	-0.02827 (0.0058) **	-0.02823 (0.0067) **	-0.02644 (0.0068) **	-0.02545 (0.0064) **	-0.02633 (0.0069) **	-0.02616 (0.0079) **	-0.02679 (0.0079) **		-0.0250 (0.0067) **	-2.15419 (0.5449) **
Number of children 0-5								-0.0298 (0.0079) **		
Number of children 6-17								-0.0095 (0.0205)		
Year (at time of test)									-0.0102 (0.0039) **	
Year squared									0.0006 (0.0003) **	
Estimation Method	LIML	LIML	LIML	LIML	LIML	LIML	LIML	LIML	LIML	LIML
Number of Observations	3787	3787	3787	3,787	3787	3787	3787	3787	3787	3787
R-squared	0.3844	0.3841	0.3848	0.3844	0.3848	0.3304	0.3803	0.3808	0.3884	0.3936
k [#]	1.005	1.007	1.006	1.005	1.007	1.013	1.005	1.005	1.008	1.005
Weak/Many Instruments Test	15.66	13.98	11.76	15.33	13.35	19.85	5.98	11.85	12.99	15.33

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

[&] Equals 1 if mother always worked since childbirth, 0.5 if she worked at least one quarter and 0 if she did not work at all (also instrumented for with same instrument list).

^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.

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

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

Table 9
Robustness with respect to mother's age at childbirth

Dependent Variable -> Log(Test Score)	(Mother's) Age Restricted Samples			Models with Age Interactions		
	24+	24-34	24-30	24+	24-34	24-30
Cumulative Child Care	-0.00585 (0.0050)	-0.00515 (0.0045)	-0.00404 (0.0058)	-0.00575 (0.0029) **	-0.00565 (0.0028) **	-0.00608 (0.0029) **
Cumulative Child Care * Mother's age ^{&}				0.00153 (0.0018)	0.00100 (0.0011)	0.00126 (0.0406)
Included Interactions (with mother's age)				Child's age AFQT missing	None	None
χ^2 -test for additional age interactions				28.44	31.82	30.61
Degrees of freedom				21	23	23
P-value				(0.1282)	(0.1040)	(0.1327)
χ^2 -test for significance of Child care + Child care*mother's age				2.03 (0.1540)	2.64 (0.1040)	2.98 (0.0845)
Estimation Method	LIML	LIML	LIML	LIML	LIML	LIML
Number of Observations	1,680	1,643	1,345	3,787	3,787	3,787
R-squared	0.3758	0.3780	0.3915	0.3872	0.3857	0.3854
k [#]	1.082	1.076	1.098	1.040	1.040	1.039
Weak/Many Instruments Test	4.25	4.49	3.85	5.68	5.72	5.72

Instruments are: All 78 policy variables, local demand conditions and interactions described in footnote in Table 5.

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

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

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

** Significant at 5%, * Significant at 10%

Table 10
Robustness with respect to State Fixed Effects

Dependent Variable -> Log(Test Score)	No State	With
	Fixed Effects	State F.E.
Cumulative Child Care	-0.00533 (0.0025) **	-0.00982 (0.0053) *
Log(Cumulative Income)	0.01062 (0.0266)	0.02863 (0.0344)
Mother's education	0.01297 (0.0030) **	0.01467 (0.0036) **
Mother's AFQT	0.00066 (0.0003) **	0.00069 (0.0003) **
Estimation Method	LIML	LIML
Number of Observations	3,787	3,787
R-squared	0.3844	0.3646
k [#]	1.005	1.004
Weak/Many Instruments Test	15.33	9.02
Joint-significance Test (State F.E.)		30.25 (0.066)

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

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

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

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

Table 11**A. Robustness with respect to the Instrument List**

Dependent Variable -> Log(Test Score)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Original set of IVs ^a	Excluding CCDF ^a	Only TL, WR and ED	Excludes TL, WR and ED	Only BEN and local demand	Original set without interactions	Only State-specific instruments ^b	Only State-specific IVs without interactions
Cumulative Child Care	-0.00522 (0.0028) *	-0.00584 (0.0028) **	-0.00758 (0.0038) **	-0.00414 (0.0024) *	-0.00437 (0.0024) *	-0.00528 (0.0028) *	-0.00623 (0.0025) **	-0.00642 (0.0026) **
Log(Cumulative Income)	0.01037 (0.0242)	0.01026 (0.0250)	0.02118 (0.0367)	0.02054 (0.0302)	-0.00294 (0.0372)	-0.00445 (0.0239)	-0.00300 (0.0246)	0.00529 (0.0361)
Mother's education	0.01276 (0.0032) **	0.01305 (0.0032) **	0.01327 (0.0034) **	0.01049 (0.0036) **	0.01208 (0.0036) **	0.01285 (0.0030) **	0.01316 (0.0034) **	0.01167 (0.0037) **
Mother's AFQT	0.00066 (0.0003) **	0.00067 (0.0003) **	0.00064 (0.0003) **	0.00056 (0.0003) *	0.00073 (0.0004) **	0.00077 (0.0003) **	0.00078 (0.0003) **	0.00073 (0.0004) *
Estimation Method	LIML	LIML	LIML	LIML	LIML	LIML	LIML	LIML
Number of Observations	3,787	3,787	3,787	3,787	3,787	3,787	3,787	3,787
R-squared	0.3847	0.3816	0.3705	0.3843	0.3860	0.3820	0.3767	0.3720
k [#]	1.040	1.039	1.036	1.004	1.003	1.008	1.026	1.006
Weak/Many Instruments Test	5.80	5.80	5.21	7.00	5.71	9.09	6.27	7.00
Number of instruments	78	75	58	27	18	26	63	25

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

^a See descriptions of instruments in Table 1. CCDF: ChildcareDevelopment 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 11A)

Dep var-> Cum. 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

R² of first stage regression with only exogenous variables=0.4765

Table12
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 (0.0029)	-0.00493 (0.0029) *	-0.00502 (0.0027) *	-0.00709 (0.0029) **	-0.00814 (0.0041) **	-0.00885 (0.0046) *	0.00465 (0.0145)
~							
Education*(Cum. Child Care)	-0.00088 (0.0005) *						
~							
AFQT*(Cum. Child Care)		-0.00011 (0.00007) *					
~							
(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 (0.0244)	0.01441 (0.0258)	0.01004 (0.0244)	0.00661 (0.0243)	0.00827 (0.0245)	0.00740 (0.0246)	0.01089 (0.0242)
Mother's Education	0.01858 (0.0042) **	0.01232 (0.0032) **	0.01288 (0.0031) **	0.01313 (0.0032) **	0.01278 (0.0031) **	0.01276 (0.0031) **	0.01251 (0.0032) **
Mother's AFQT score	0.00066 (0.0003) **	0.00171 (0.0007) **	0.00065 (0.0003) **	0.00068 (0.0003) **	0.00072 (0.0003) **	0.00076 (0.0003) **	0.00067 (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)	

Education=Education-Education, where Education is the mean (same for number of children and mother's AFQT score).

Instruments are: All 78 policy variables, local demand conditions and interactions described in footnote in 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 13
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 (5.873)	0.00067 (0.0008)	-0.00643 (0.0029) **			-0.00638 (0.0031) **
Informal child care * I[mother college graduate]						-0.00032 (0.0023)
Relatives	5.0077 (5.736)			0.00034 (0.0008)	-0.00751 (0.0029) **	
Nonrelatives	1.1454 (3.355)			0.00164 (0.0011)	0.00683 (0.0069)	
Cumulative Formal child care (Daycare, Nursery, Pre-K, Other)	1.2229 (3.055)	0.00307 (0.0011) **	0.00302 (0.0066)	0.00309 (0.0011) **	0.00472 (0.0066)	0.00831 (0.0129)
Formal child care * I[mother college graduate]						-0.00703 (0.0101)
Log(Cumulative Income)	3.6332 (0.730)	-0.00343 (0.0057)	0.00719 (0.0233)	-0.00331 (0.0057)	0.00883 (0.0230)	0.01094 (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 all 78 policy variables, local demand conditions and interactions described in footnote in 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).

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 14
Effect of Child Care on Non-Cognitive Outcomes

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

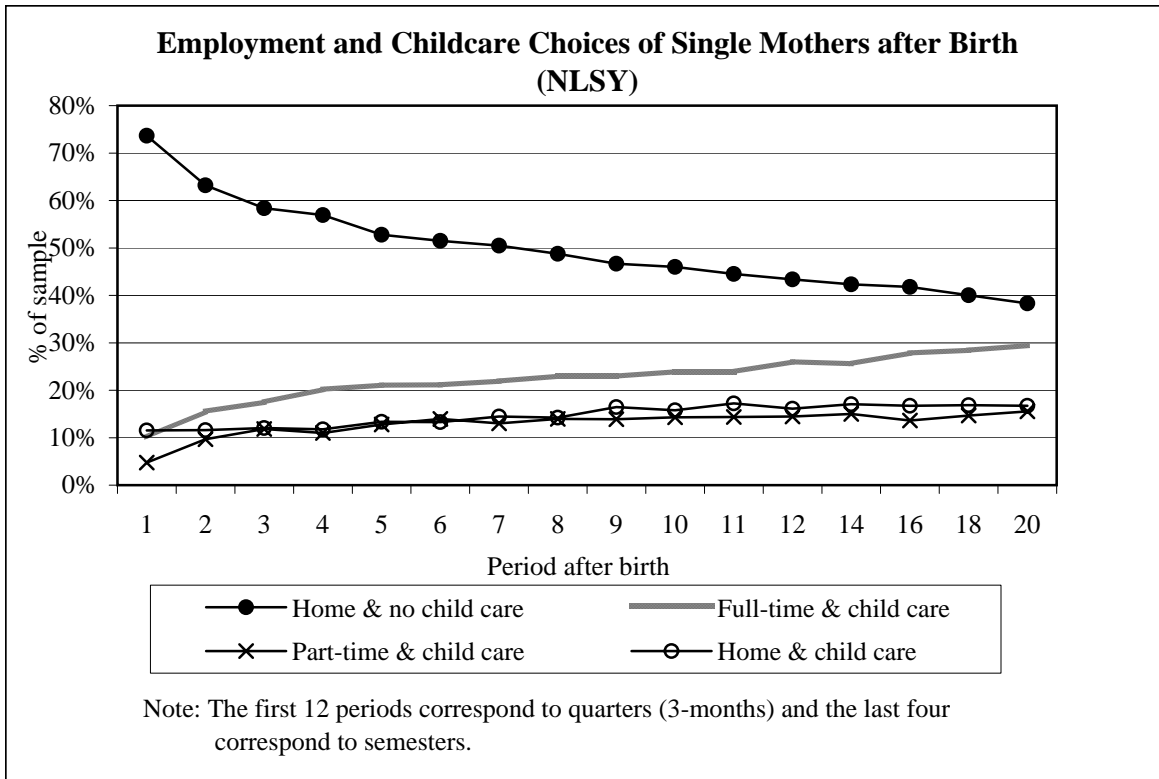
^a Excluded category: widowed.

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



Home & no child care Full-time & child care Part-time & child care Home & child care

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 (Test and age in column heading)	0.00819 (0.0041) **	0.01574 (0.0035) **	0.00633 (0.0046)	0.01627 (0.0044) **	0.00960 (0.0048) **	0.02092 (0.0045) **	0.01908 (0.0049) **	0.03166 (0.0045) **	0.02493 (0.0056) **	0.03744 (0.0055) **
Age of completed education measure#	0.16449 (0.1563)	0.06336 (0.1575)	0.69629 (0.0752) **	0.68394 (0.0717) **	0.69097 (0.0758) **	0.66007 (0.0723) **	0.45305 (0.0438) **	0.41675 (0.0409) **	0.45629 (0.0439) **	0.40288 (0.0411) **
Highest grade completed by mother	0.09231 (0.0403) **		0.05216 (0.0348) *		0.04901 (0.0343)		0.09646 (0.0270) **		0.10179 (0.0268) **	
Highest grade completed by father	0.02147 (0.0083) **		0.02069 (0.0076) **		0.01948 (0.0075) **		0.00833 (0.0064)		0.01065 (0.0064) *	
Number of siblings	-0.14160 (0.0586) **		-0.14066 (0.0543) *		-0.12912 (0.0535) **		-0.08883 (0.0428) *		-0.08942 (0.0424) **	
Birthorder	-0.11146 (0.0979)		-0.13111 (0.0957)		-0.09435 (0.0946)		-0.11223 (0.0754)		-0.07853 (0.0751)	
Race (1=Non-white)	0.06958 (0.1751)		0.08739 (0.1523)		0.06939 (0.1496)		-0.06182 (0.1258)		-0.21639 (0.1243) *	
Gender (1=Male)	-0.36024 (0.1380) **		-0.42114 (0.1236) **		-0.42716 (0.1228) **		-0.39478 (0.1011) **		-0.37505 (0.1008) **	
Mother's age at child's birth	-0.03878 (0.0387)		-0.01219 (0.0336)		-0.02523 (0.0331)		0.02586 (0.0282)		0.03390 (0.0280)	
Mother's AFQT score	0.00389 (0.0038)		0.00378 (0.0033)		0.00450 (0.0033)		0.00128 (0.0029)		-0.00030 (0.0028)	
Constant	7.2531 (3.1866) **	8.3078 (2.9599) **	-2.8778 (1.8248)	-3.8171 (1.4088) **	-2.9770 (1.7977)	-4.0097 (1.3892) **	-0.6925 (1.2501)	-0.1622 (-0.1622) **	-1.5869 (1.2644)	-0.6049 (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) **
Estimation	Probit
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

Child's Age	PPVT			PIAT - Math		PIAT-Reading	
	3	4	5	5	6	5	6
Log(test score) in our sample	4.367 (0.191)	4.2689 (0.295)	4.402 (0.239)	4.539 (0.152)	4.543 (0.128)	4.633 (0.152)	4.606 (0.095)
Test Scores in our sample	80.263 (14.952)	74.334 (19.512)	83.767 (17.504)	94.719 (14.329)	94.802 (11.727)	104.089 (15.319)	100.585 (9.462)
Non-whites	78.007 (14.169)	70.836 (17.958)	82.135 (16.889)	93.836 (14.289)	94.247 (11.685)	103.358 (15.454)	100.482 (9.269)
Whites	92.167 (13.348)	89.299 (18.885)	93.852 (18.001)	99.576 (13.634)	97.657 (11.578)	108.100 (13.970)	101.112 (10.422)
Maternal education (12 yrs+)	82.820 (14.369)	78.748 (18.917)	88.743 (17.648)	97.084 (14.178)	96.823 (11.663)	106.755 (15.131)	102.265 (9.425)
Maternal education (<12 yrs)	76.301 (15.025)	68.748 (18.847)	79.508 (16.245)	91.767 (13.991)	92.751 (11.449)	100.697 (14.909)	98.847 (9.197)
Male	79.753 (14.664)	72.242 (20.048)	83.035 (18.143)	93.726 (14.307)	93.710 (12.292)	102.557 (15.563)	99.232 (9.404)
Female	80.707 (15.225)	76.299 (18.820)	84.569 (16.783)	95.739 (14.305)	95.827 (11.091)	105.685 (14.922)	101.838 (9.357)
No. of observations	339	512	438	598	663	584	653

PPVT: Peabody Picture Vocabulary Test

PIAT: Peabody Individual Achievement Test

Standard errors in parenthesis.

Appendix 4

Let S_3, S_4, S_5 and S_6 be the child's test scores at ages 3 through 6, respectively.¹ For example, S_3 can be the PPVT score at age 3.² In addition, let Y_3, Y_4 and Y_5 represent the endogenous variables that appear in the test score equation (10) in year 3, 4 and 5 after childbirth respectively. For example, Y_5 would include cumulative child care use up through age 5. Finally, let R_1, R_2, \dots, R_5 represent vectors of instruments that are relevant for the mother's decisions in years 1 through 5 after the birth of the child. Thus, for example, R_5 would include welfare policy rules operative in the state of residence of the mother in year 5,³ and Y_5 is potentially influenced by R_1 through R_5 . The first stage regressions in the 2SLS procedure will thus look like:

$$\begin{aligned} Y_{3i} &= \alpha_0 + \alpha_1 R_{1i} + \alpha_2 R_{2i} + \alpha_3 R_{3i} + \alpha_4 \cdot 0 + \alpha_5 \cdot 0 + \underline{\alpha_6} X_{3i} + \varepsilon_i \\ Y_{4i} &= \alpha_0 + \alpha_1 R_{1i} + \alpha_2 R_{2i} + \alpha_3 R_{3i} + \alpha_4 R_{4i} + \alpha_5 \cdot 0 + \underline{\alpha_7} X_{4i} + \varepsilon_i \\ Y_{5i} &= \alpha_0 + \alpha_1 R_{1i} + \alpha_2 R_{2i} + \alpha_3 R_{3i} + \alpha_4 R_{4i} + \alpha_5 R_{5i} + \underline{\alpha_8} X_{5i} + \varepsilon_i \end{aligned} \quad (\text{A4.1})$$

where X_{ti} is a vector of exogenous characteristics of mothers and children that include all variables described in Table 4, and $\underline{\alpha_6}$ is an associated parameter vector. Notice that R_1 through R_5 all enter the equation for Y_t .

From A4.1, we obtain the fitted values \hat{Y}_{ti} for $t=3, 4, 5$.

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

$$\begin{aligned} S_{3i} &= \beta_0 + \beta_1 \hat{Y}_{3i} + \underline{\beta_2} X_{3i} + \xi_i \\ S_{4i} &= \beta_0 + \beta_1 \hat{Y}_{4i} + \underline{\beta_2} X_{4i} + \xi_i \\ S_{5i} &= \beta_0 + \beta_1 \hat{Y}_{5i} + \underline{\beta_2} X_{5i} + \xi_i \\ S_{6i} &= \beta_0 + \beta_1 \hat{Y}_{5i} + \underline{\beta_2} X_{6i} + \xi_i \end{aligned} \quad (\text{A4.2})$$

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

In the baseline specification, in order to avoid proliferation of parameters, we estimate a constrained version of the 1st stage regressions A4.1 where we assume the effects of the instruments on the endogenous variable Y_{ti} are the same in every year after birth, i.e., $\alpha_1 = \alpha_2 = \dots = \alpha_5$.⁴ Also, we constrain $\underline{\alpha_6}$, $\underline{\alpha_7}$ and $\underline{\alpha_8}$ to differ only in that a subset of the elements of X are interacted with child age.⁵

¹ Recall that cognitive ability test scores are available as early as age 3 in the NLSY.

² Since we have quarterly data, a test score at 3 literally means a test score in the 12th quarter after the birth of the child.

³ As well as interactions of the policy rules with mother's education and AFQT.

⁴ In one robustness test (Table 14 column 7) we interact all the instruments with child age, but it makes little difference.

⁵ These are workbef, EXPBEF, urban and age of mother. Other interactions were insignificant.