*2010 International Conference on Business and Economics Research*
*vol.1 (2011) © (2011) IACSIT Press, Kuala Lumpur, Malaysia*

# Is There a Trade-off between Child Quantity and Child Quality? A Case Study of Children in Rural Terengganu, Malaysia

Nor Azam Abdul-Razak, Roslan Abdul-Hakim and Russayani Ismail
Department of Economics, College of Arts and Sciences
Universiti Utara Malaysia, UUM Sintok
Kedah, Malaysia
azam@uum.edu.my

***Abstract*—The objective of this paper is to examine whether the hypothesis of the child quantity-quality trade-off developed by (3] and [4] is borne out by data from a developing country. In particular, this paper tests whether child quantity has a negative impact on child quality (i.e. child's education) using a sample of children data from a recent survey of 2500 households in rural areas in Terengganu, Malaysia. Like previous empirical studies, this study employs an instrumental variable method in a system of two equations, where child quantity is instrumented by an exogenous variation in fertility, namely gender sameness of the first two-born children. In the baseline estimation as well as in a series of robustness check, our key findings are that there is a positive yet insignificant impact of child quantity on child quality. Accordingly, we take these results as weak evidence against the child quantity-quality trade-off.**

***Keywords-child quantity; child quality; trade-off; fertility; human capital; gender sameness.***

## I. INTRODUCTION

Human capital is vital in promoting economic growth and development of a country. High quality human capital is usually associated with high productivity and innovativeness which is the basic ingredient for achieving rapid economic growth. The question is therefore how we promote quality human capital so as to achieve higher growth. In this respect, it appears that the decision of households with regard to fertility has an important consequence on the future stock of human capital of a nation. This entails an investigation on the determinants of the level of human capital of children. In this regard, economists have long observed that large families tend to have relatively uneducated children while small families tend to have relatively educated children. What could have accounted for the differences in the behavior of these families?

In a pioneering work, [3] and [4] hypothesize that as household income increases, the marginal costs of child quantity and child quality are expected to increase as well. To illustrate, consider two households, A and B, each of whom has the same number of children (say, 3) but A's children are more educated. Then, adding another child is more costly for A. In other words, the marginal cost of child quantity is higher for A. Now consider another two households, X and Z, each of whom has children of the same level of education (say, 16 years of education) but X has more children than Z. Then, adding another year of education is more costly for X. In other words, the marginal cost of child quality is higher for X.

Given a fixed budget constraint, the rising marginal costs of child quantity and child quality imply that a given household has to choose between having either more children or educated children; hence, the trade-off between child quantity and child quality.

Following [3] and [4], numerous empirical studies have been conducted to examine this postulated trade-off by investigating whether child quantity exerts a negative impact on child quality. In most of these early studies, however, it was found that child quantity has a negative effect on child quality, thus lending support to the theory (see [6] for a survey of these studies). In recent years, however, some influential studies have found that there is lack of evidence of the child quantity-quality trade-off (see [5] and [2]). The discrepancy between these findings, some argue, could be attributed to the problem of endogeneity. In particular, older studies treat child quantity as an exogenous variable whereas newer studies treat child quantity as an endogenous variable. (An important exception is an early study by [10].) From the theoretical point of view, child quantity is an endogenous variable since households choose both the quantity and quality of their children. As such, the results of these newer studies should be more trustworthy.

In another recent study, however, [9] find that there is a trade-off even after tackling the problem of endogeneity. According to these scholars, the discrepancy between their results and those of [5] and [2] could be attributed to the extent of public provision of education. In the case of China, they argue, there is a poor provision of public education, thus parents have to bear the bulk of education costs of their children. Consequently, child quantity-quality trade-off is prevalent in the case of China. In light of this argument, this paper revisits the issue by specifically examining whether [9]'s argument holds in the case of Malaysia. Since education is highly subsidized in Malaysia, we expect that there is little or no trade-off in the country.

## II. MODEL SPECIFICATION

Following the literature, we specify child quality (measured by child's educational attainment) as a function of child quantity (or fertility) and a host of household characteristics as control variables. In view of its

endogeneity, child quantity needs to be instrumented by an appropriate instrumental variable. A natural candidate for the instrument, as suggested by the empirical literature, is some exogenous variation in fertility as (i.e. a variation in fertility that is exogenous to the choice of child quality but is correlated with the choice of child quantity).

Hitherto, two frequently used exogenous variations in fertility are the occurrence of twins (defined as a dummy variable which is equal to 1 if the $n^{th}$ birth is a twin and 0 otherwise) and gender sameness (defined as a dummy variable which is equal to 1 if the first $n^{th}$ children are of the same gender and 0 otherwise). The basis for their choice is two-fold. First, it is plausible to assume (and even confirm) that each of the candidate instruments is correlated with child quantity. Second, it is reasonable to argue that each of the candidate instruments is uncorrelated with child quality (except through child quantity) because it is unlikely that parents would underinvest in the education of their children simply because they are endowed with twin children or same-gender children.

In the studies conducted by [5] and [9], twins have been employed as the instrument. It is imperative therefore that the present study employ the same variable. Since the twin data are not available in our data set, this study opts for gender sameness as the instrument for child quantity. This variable has been employed by [1] and [7], among others. Accordingly, our model can be specified as follows:

$$Educ_i = \beta_1 + \beta_2 Kids_i + \mathbf{x_i}'\boldsymbol{\gamma} + u_i \quad (1)$$

$$Kids_i = \alpha_1 + \alpha_2 Sameness_i + \mathbf{x_i}'\boldsymbol{\delta} + v_i \quad (2)$$

where $Educ_i$ is child $i$'s educational attainment (measured by the number of years of schooling), *Kids* is child quantity (i.e. the total number of children in a household), *Sameness* is a dummy variable for gender sameness (which is equal to 1 if the first two children are of the same gender and 0 otherwise), and **x** is a vector of control variables which include child's characteristics (i.e. gender, birth order, and age) and parents' characteristics (i.e. parents' age and parents' education).

A priori, we expect the coefficient of *Kids* to be negative to reflect the adverse relationship between child quantity and child quality. In contrast, we expect the coefficient of *Sameness* to be positive to reflect parents' preferences for children of mixed gender.

## III. DATA

The data in this study are obtained from an interview-based survey conducted on a sample of 2500 rural households in Terengganu, Malaysia in May 2009. Since the unit of analysis is children (instead of households), we extract the children data from these households. From the 2500 households, the total number of children is 12,321 persons. This means that, on average, there are about 5 children per household.

One of the remarkable features of the empirical work on the child quantity-quality trade-off is that the sample is subject to a number of restrictions. First, the sample needs to be restricted to school-aged children who are currently residing with their parents. This is usually accomplished by restricting the sample to children who are aged 5-17 or 6-17 (see [7] and [9], respectively). In the case of Malaysia, children begin schooling at the age of 7, and they are not so eager to leave their parents' home once they graduate from high school. Hence, the children's age range can be expanded to, say, 7-20. Doing so reduces our sample size to 4760 observations (i.e. a reduction by 7561 observations).

Second, depending on the way the instrument for child quantity is defined, the sample needs to be restricted further. If gender sameness of the first two-born children is chosen as the instrument, then the sample needs to be restricted to households who have at least 3 children. Doing so decreases our sample size to 4569 observations (i.e. a decline by 191 observations). Third, given the instrument, too, the unit of analysis needs to be confined to the first two-born children. Doing so results in a huge drop in our sample size (i.e. by 3374 observations), leaving us with 1195 observations (see [5] and [7] for the second and third sample restrictions).

Finally, the data for some of the variables of interest in this curtailed sample are either not available or suspicious. These missing and dubious values, in turn, shrink the sample size further by 310 observations. Hence, we end up with 885 observations. This figure corresponds to 430 households, all of whom are characterized by dual parents, 311 of whom (or 72%) are characterized by dual income earners, and 230 of whom (or 53%) are endowed with the first two-born children of the same gender.

Given this substantially reduced sample size, the summary statistics of the key variables are as follows: *Educ* (i.e. children's educational attainment) ranges from 0 to 17 years with the average of about 7.7 years, *Kids* (i.e. the number of children) ranges from 3 to 10 children with the average of about 4 children, age of the children ranges from 7 to 20 years with the average of about 14 years, 458 of the children (or 52%) are males, father's age ranges from 26 to 75 years with the average of about 43 years, mother's age ranges from 24 to 62 years with the average of about 38 years, father's educational attainment ranges from 0 to 18 years with the average of about 9 years, and mother's educational attainment ranges from 0 to 19 years with the average of about 9 years, and *HHInc* (i.e. household income) ranges from RM250 to RM11250 with the average of about RM1763 (see Table 1 for a more detailed summary statistics of all of these variables).

TABLE I. DESCRIPTIVE STATISTICS OF THE VARIABLES (N = 885)

| | Min. | Max. | Mean | Std. Dev. |
|---|---|---|---|---|
| Educ (years) | 0 | 17 | 7.66 | 3.67 |
| Kids (quantity) | 3 | 10 | 4.34 | 1.31 |
| Child's Age (years) | 7 | 20 | 14.14 | 3.82 |
| Father's Age (years) | 26 | 75 | 42.87 | 6.33 |
| Mother's Age (years) | 24 | 62 | 37.73 | 5.11 |
| Father's Educ (years) | 0 | 18 | 9.34 | 3.56 |
| Mother's Educ (years) | 0 | 19 | 9.25 | 3.25 |

| | | | | |
|---|---|---|---|---|
| HHInc (RM) | 250 | 11250 | 1762.88 | 1571.45 |

## IV. Estimation Results

Given the necessary data for a sample of 885 children in Terengganu, Malaysia, we estimate (1) by the instrumental variable (IV) method, where *Kids* (i.e. child quantity) is instrumented by *Sameness* (i.e. gender sameness of the first two-born children), and (2) by the ordinary least squares (OLS) method. In other words, (2) serves as the first-stage regression and (1) the second-stage regression.

We begin by running the first-stage regression corresponding to (2) and present the estimation results in Table (1a). Column 1 shows that the coefficient of *Sameness* enters with the unexpected negative sign but it is significant at the 5% level. The negative sign of the coefficient carries the counterintuitive implication that households with the first two-born children of the same gender have a tendency to stop reproduction. Nonetheless, since the coefficient is statistically significant, we proceed with the second-stage regression corresponding to (1) and present the results in Column (1b). We observe that the coefficient of *Kids* enters with the unexpected positive sign but it is insignificant even at the 10% level. We take these results as mild evidence against the theory of child quantity-quality trade-off.

It could be argued that parents behave differently with respect to the arrival of the first two-born boys than the first two-born girls. If parents prefer boys to girls, as argued by some scholars (see, for example, [8]), then they are more inclined to consider having another child if their first two-born children are girls rather than boys. Hence, combining these two cohorts of children might undermine or even distort the instrument's impact on child quantity.

To accommodate this line of reasoning, we repeat the analysis with gender sameness of the first two-born boys, *BB*, as the instrument for child quantity. As before, we run the first-stage regression and report the results in Column (2a). We observe that the coefficient of *BB* enters with the unexpected negative sign but it is significant at the 1% level. In the second-stage regression, as shown in Column (2b), we find that the coefficient of *Kids* enters with the unexpected positive sign and is significant at the 10% level. We take these results as evidence against the theory of child quantity-quality trade-off.

For completeness, we repeat the analysis once again with gender sameness of the first two-born girls, *GG*, as the instrument for child quantity. The results of the first-stage regression indicate that the coefficient of *GG* enters with the unexpected negative sign and is insignificant even at the 10% level. (Due to space constraints, these results are not tabulated.) The statistical insignificance of *GG* indicates that the variable is a weak instrument, thereby precluding us from proceeding with the second-stage regression.

TABLE II. Baseline Estimation Results (N = 885)

| | (1a) | (1b) | (2a) | (2b) |
|---|---|---|---|---|
| Dependent Variable | Kids | Educ | Kids | Educ |
| Constant | 3.235*** (9.25) | −8.600*** (−3.90) | 3.118*** (8.99) | −10.358*** (−5.10) |
| Kids | – | 0.553 (0.81) | – | 1.111* (1.80) |
| Sameness | −0.163** (−2.07) | – | – | – |
| BB | – | – | −0.301*** (-2.77) | – |
| Child's Gender | −0.122 (−1.54) | −0.032 (−0.23) | 0.037 (0.38) | 0.038 (0.24) |
| Second Child | 0.427*** (5.18) | −0.233 (−0.74) | 0 .432*** (5.25) | −0.472 (−1.58) |
| Child's Age | 0.187*** (13.39) | 0.741*** (5.74) | 0.186*** (13.33) | 0.636*** (5.40) |
| Father's Age | 0.009 (1.00) | 0.024* (1.65) | 0.010 (1.09) | 0.019 (1.13) |
| Mother's Age | −0.061*** (−4.97) | 0.046 (1.01) | −0.061* (−4.95) | 0.080* (1.85) |
| Father's Educ | 0.026* (1.92) | 0.036 (1.40) | 0.026* (1.96) | 0.022 (0.78) |
| Mother's Educ | 0.006 (0.43) | 0.045** (2.15) | 0.005 (0.37) | 0.042* (1.66) |
| Adj. $R^2$ | 0.20 | 0.80 | 0.21 | 0.71 |

Note: Estimation is done by the IV method; columns (a) and (b) report the results from first- and second-stage regressions, respectively. The figures in parentheses are t-values; ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

It could be argued that low-income households are especially concerned about their own welfare during their old age. Given their tight budget constraints as well as more costly and risky investment in child quality, an investment in child quantity appears to be a better means of providing these low-income households with a social safety net during their old-age period. If so, then an increase in household income is more likely to induce the low income-households to have more children (instead of educated children). This implies that the trade-off between child quantity and child quality is unlikely to be applicable to low-income households.

Given this argument, a dummy variable representing low-income households needs to be introduced into our model. Recall from Table 1 that the average household income is about RM1763 per month. Thus, setting RM1500 as the cut-off point between low- and high-income households, a dummy variable defined as *LT1500* is introduced for those who earn less than RM1500 per month. The empirical analysis is repeated with the inclusion of this new variable and the results are documented in Table 3. In the first-stage regression, we see that the coefficient of *Sameness* continues to enter with the negative sign and is significant at the 5% level [see Column (1a)]. Likewise, the coefficient of *Kids* in the second-stage regression continues to enter with the positive sign and is insignificant even at the 10% level [see Column (1b)].

TABLE III. Estimation Results with the Inclusion of LT1500 as another Control Variable (N = 885)

| | (1a) | (1b) | (2a) | (2b) |
|---|---|---|---|---|
| Dependent Variable | Kids | Educ | Kids | Educ |
| Constant | 3.433*** (9.09) | −8.964*** (−3.98) | 3.318*** (8.86) | −10.821*** (−5.23) |
| Kids | – | 0.497 (0.76) | – | 1.054* (1.78) |
| Sameness | −0.166** (−2.12) | – | – | – |
| BB | – | – | −0.306*** (−2.82) | – |
| LT1500 | −0.122 (−1.38) | 0.336** (2.32) | −0.125 (−1.41) | 0.401** (2.47) |
| Child's Gender | −0.122 (−1.55) | −0.037 (−0.27) | 0.039 (0.40) | 0.033 (0.22) |
| Second Child | 0.422*** (5.11) | −0.194 (−0.64) | 0.426*** (5.18) | −0.429 (−1.51) |

| | | | | |
|---|---|---|---|---|
| Child's Age | 0.184*** (13.15) | 0.757*** (6.15) | 0.183*** (13.08) | 0.654*** (5.85) |
| Father's Age | 0.010 (1.07) | 0.023 (1.56) | 0.011 (1.17) | 0.017 (1.04) |
| Mother's Age | −0.062*** (−5.06) | 0.046 (1.03) | −0.062*** (−5.05) | 0.080* (1.91) |
| Father's Educ | 0.022 (1.57) | 0.050** (2.06) | 0.022 (1.59) | 0.037 (1.41) |
| Mother's Educ | 0.003 (0.22) | 0.054*** (2.63) | 0.002 (0.16) | 0.052** (2.13) |
| Adj. $R^2$ | 0.20 | 0.80 | 0.21 | 0.72 |

Note: Estimation is done by the IV method; columns (a) and (b) report the results from first- and second-stage regressions, respectively. The figures in parentheses are t-values; ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively..

As before, we repeat the analysis with *BB* as the instrument for child quantity in lieu of *Sameness*. As shown in Columns (2a) and (2b), the coefficients of interest (i.e. those of *BB* and *Kids* in the first- and second-stage regressions, respectively) enter with the same signs and significance levels as their counterparts in the baseline analysis. Finally, the analysis is repeated with *GG* as the instrument for child quantity. As before, we find that the coefficient of *GG* enters with the negative sign and is insignificant. (Due to space constraints, these results are not tabulated.) Taken together, all of these results suggest that conditioning our analysis to low-income households does not affect the trade-off (or lack of it) between child quantity and child quality.

It could be argued that low-income households are not particularly concerned about the gender of the first two-born children. In order to provide them for a social safety net during their old age, these low-income households might continue to have more children regardless of the gender of the first two-born children. For non-poor households, however, the decision to have more children might depend upon the gender of the first two-born; i.e. they are more likely to add a third child if the first two-born are of the same gender. If so, then gender sameness is unlikely to be applicable to low-income households.

Given this conjecture, the dummy *LT1500* qualifies as an additional instrument for child quality. The empirical analysis is repeated with the inclusion of this additional dummy in the first-stage regression and the results are reported in Table 4. As shown in Column (1a), the coefficient of *Sameness* enters with the same negative sign and is significant at the 5% level; the coefficient of *LT1500* enters with the negative sign and is insignificant. However, since both coefficients are jointly significant at the 5% level, we proceed with the second-stage regression. Unlike the previous findings, here we observe that the coefficient of *Kids* enters with the expected negative sign albeit it is insignificant [see Column (1b)].

TABLE IV. ESTIMATION RESULTS WITH THE

Inclusion of LT1500 as another Instrument (N = 885)

| | (1a) | (1b) | (2a) | (2b) |
|---|---|---|---|---|
| Dependent Variable | Kids | Educ | Kids | Educ |
| Constant | 3.433*** (9.09) | −5.893*** (−3.31) | 3.318*** (8.86) | −8.234*** (5.58) |
| Kids | – | −0.306 (−0.56) | – | 0.437 (0.99) |
| Sameness | −0.166** (−2.12) | – | – | – |
| BB | – | – | −0.306*** (−2.82) | – |
| LT1500 | −0.122 (−1.38) | – | −0.125 (−1.41) | – |
| Child's Gender | −0.122 (−1.55) | −0.140 (−1.10) | 0.039 (0.40) | −0.047 (−0.38) |
| Second Child | 0.422*** (5.11) | 0.135 (0.52) | 0.426*** (5.18) | −0.183 (−0.83) |
| Child's Age | 0.184*** (13.15) | 0.902*** (8.69) | 0.183*** (13.08) | 0.762*** (8.95) |
| Father's Age | 0.010 (1.07) | 0.032** (2.35) | 0.011 (1.17) | 0.025* (1.87) |
| Mother's Age | −0.062*** (−5.06) | −0.007 (−0.18) | −0.062*** (−5.05) | 0.039 (1.21) |
| Father's Educ | 0.022 (1.57) | 0.059** (2.55) | 0.022 (1.59) | 0.039* (1.81) |
| Mother's Educ | 0.033 (0.22) | 0.051** (2.53) | 0.002 (0.16) | 0.046** (2.28) |
| Adj. $R^2$ | 0.20 | 0.81 | 0.21 | 0.81 |

Note: Estimation is done by the IV method; columns (a) and (b) report the results from first- and second-stage regressions, respectively. The figures in parentheses are t-values; ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Repeating the analysis with *BB* in place of *Sameness*, we see that the coefficient of *BB* enters with the usual negative sign and is significant at the 1% level; the coefficient of *LT1500* enters with the negative sign and is insignificant [see Column (2a)]. Again, since both coefficients are jointly significant at the 1% level, we proceed with the second-stage regression. In this case, the coefficient of *Kids* enters with the usual positive sign and it is insignificant [see Column (2b)]. Finally, the analysis is repeated with *GG* in lieu of *Sameness*. As before, the coefficient of *GG* enters with the negative sign and is insignificant; so does the coefficient of *LT15*. (Due to space constraints, these results are not tabulated.) Since both coefficients are jointly insignificant, however, the second-stage regression is uncalled for. Here the evidence is mixed: when *Sameness* is used as the instrument, there is weak evidence in favor of the child quantity-quality trade-off; when *BB* is used as the instrument, there is weak evidence against the trade-off.

Overall, our empirical results indicate that there is a positive yet insignificant impact of child quantity on child quality, suggesting that there is mild evidence against the theory of child quantity-quality trade-off. In all fairness, this weak evidence implies that it is unclear whether a growing Malaysian economy is expected to bring about a decline in its population and a rise in the stock of human capital.

## V. CONCLUSION

In this paper, we revisit the issue of child quantity-quality trade-off in the context of a rural area in a developing country, Malaysia. Our baseline findings indicate that the estimated coefficient of child quantity is positive but insignificant even at the 10% level. It appears that these results are broadly robust to the conditioning of our sample on low-income households, *LT1500*, either as an additional control variable or as an additional instrument. Accordingly, we take all of these results as weak evidence against the theory of the trade-off between child quantity and child quality. Our results appear to be consistent with those of [5]

and [2] but inconsistent with [9]. Nevertheless, we agree with [9]'s argument that the trade-off between child quantity and child quality depends on the extent of public support for education. In the case of Malaysia, primary and secondary education is basically provided for free, and tertiary education is highly subsidized.

## REFERENCES


[1] J. Angrist and W. Evans, "Children and their parent's labor supply: Evidence from exogenous variation in family size," American Economic Review, vol. 88, 1988, pp. 450–477.

[2] J. Angrist, V. Lavy, and V. Schlosser, "New evidence on the causal link between the quantity and quality of children," NBER Working Paper 11835, 2005.

[3] G. Becker and H. Lewis, "On the interaction between the quantity and quality of children," Journal of Political Economy, vol. 81, 1973, pp. S279–S288.

[4] G. Becker and N. Tomes, "Child endowments and the quantity and quality of children," Journal of Political Economy, vol. 84, 1976, pp. S143–S162.

[5] S. Black, P. Devereux, and K. Salvanes, "The more the merrier? The effect of family size and birth order on children's education," Quarterly Journal of Economics, vol. 120, 2005, pp. 669–700.

[6] J. Blake, Family size and achievement, Berkeley and Los Angeles, CA: University of California Press, 1989.

[7] D. Conley and R. Glauber, "Parental educational investment and children's academic risk: Estimates of the impact of sibship size and birth order from exogenous variation in fertility," Journal of Human Resources, vol. 41, 2006, pp. 722–737.

[8] J. Lee, "Sibling size and investment in children's education: An Asian instrument," Journal of Population Economics, vol. 21, 2008, pp. 855–875.

[9] H. Li, J. Zhang, and Y. Zhu, "The quantity-quality trade-off of children in a developing country: Identification using Chinese twins," Demography, vol. 45, 2008, pp. 223–243.

[10] M. Rosenzweig and K. Wolpin, "Testing the quantity-quality fertility model: The use of twins as a natural experiment," Econometrica, vol. 48, 1980, pp. 227–240.