Michael Moffatt
Student ID: 5566623
moffam97@zedat.fu-berlin.de
03. July, 2024

Sibling Composition and Educational Inequality: The German Context

Abstract:

Family Size and sibling configuration was once a prolific line of research in sociology of family and education. Pioneering scholars in this niche field once consistently found a negative correlation between number of siblings and various measures of educational attainment and intelligence. In recent years, however, interpretations of causality have been brought into question due to the endogenous nature of family size. One common antidote to this is instrumental variable (IV) analysis. While most instruments have supported skepticism for causality, Jaeger (2006) and Radyakin (2007) independently found that instrumenting family size with each parent's number of siblings and age at first birth not only supports causality, but infers that standard OLS estimates underestimate it. Using the same data source and nearly identical methods and measurements, the current thesis replicates and expands the econometric analysis of Radyakin (2007). While results at first agree with the promising use of these instruments, further diagnostic tests ultimately disprove their validity.

Table of Contents

List of tables and figures:

Chapter one: Introduction

1.1 Research Topic and Question

The empirical goal of this thesis is to explore the relationship between sibling composition and educational attainment. Once a robust finding, number of siblings (hereafter referred to as sibship size) was a persistent predictor of lower educational attainment. Two mainstream theories explaining this phenomenon were resource dilution theory (RDT) and confluence theory (CT). The former posits that the limited financial, time-based, and cultural resources of parents thin out as more children are born. The latter theory attributes this relationship to the intellectual environment in which children are raised, arguing that the average intellectual milieu of a family is muddied by the birth of another child. However, more recent scrutiny has deemed these findings to be spurious at best—a methodological illusion caused by selection bias. From this skepticism emerged a third theory which posits that less educated parents are more likely to create larger families. This thesis aims to understand which of these three theories best explains the proposed trade-off between child-quantity and child-quality. In particular, the research questions proposed are as follows: (1) to which extent does sibship size predict educational attainment when controlling for selection bias and (2) which mechanisms best explain this relationship? To answer these questions, I replicate the methods of Radyakin (2007), applying instrumental variable analysis as a means of controlling for the selection bias proposed by skeptics of sibship size literature—namely Guo and VanWey (1999).

1.2 Relevance of the Topic

1.2.1 Theoretical Relevance

The role of the family in educational attainment has long been an interest of sociologists. Pierre Bourdieu (1973) co-opted Marx's conception of social and cultural reproduction, arguing that since legitimate culture is constructed by the dominant class, education is biased against individuals born into working class families, whose inherited habitus and cultural capital is deemed less valuable in the spheres of education. It can therefore be expected that working class families are less likely to obtain higher levels of education if the households in which they are nested are ill equipped at preparing them for educational institutions not aimed at creating equal opportunity, but instead at replicating social inequalities onto the next generation. Bourdieu (1973) cements this argument by proposing that "the action of school, whose effect is unequal…among children of different social classes, and whose success varies considerably among those it does have an effect, tends to reinforce and consecrate by its sanctions and initial inequalities” (pp. 69).

Sociologists, economists, and demographers alike have attempted to demonstrate empirical evidence for cultural and social reproduction in various iterations since its development. The goal of social mobility research, for instance, is to determine the causality of intergenerational transmission of education, income, and occupational class. While cross-sectional evidence for intergenerational transfers in education are richly available, unobservable confounding neighborhood and genetic influences may hide selection processes which bias these estimates (Black and Deveraux, 2005). Proving causality thus comes with a swath of methodological challenges of disentangling factors of nature from that of nurture. In their review of social mobility research, Black and Deveraux (2011) identify, evaluate, and critique three broad approaches in this endeavor: twin correlations, sampling adoptees, and educational policy changes as instrumental variables. Since these approaches only indirectly measure causality with quasi-experiments and respectively hold much room for error, estimates of intergenerational transmission vary widely both between and within these three approaches. In response to this problem, Fleury and Giles (2018) apply a meta-regression analysis of both published and unpublished studies implementing all three approaches. They find a true causal effect size of 0.15, demonstrating evidence for intergenerational transmission even controlling for an evident publication bias (pp. 169). Causality implies that beyond genetic inheritance of cognitive ability, parental education changes the social environment in which children are raised. In other words, as argued by Black and Deveraux (2005), the story of causation is that obtaining a higher education “makes you a different kind of parent” which in turn shapes the outcome of their offspring (pp. 2).

Evidence for intergenerational transmission of human capital by itself however, is not indicative of the process of cultural reproduction, since transmission of cultural capital is a key mechanism untested in conventional social mobility research. Jæger and Breen (2016) argued that evidence for transmission of cultural capital can be divided into three categories. First, a handful of studies present evidence for cultural transmission from parent to child (Kallunki and Purhonen, 2017;

Kraaykamp and Van Eijck, 2010; Georg, 2004). Second, a larger majority of literature has found evidence for a direct effect of cultural capital on educational performance and attainment (DiMaggio and Mohr, 1985; Sullivan, 2001; De Graaf and De Graaf, 2002; Tramonte and Willms, 2010). Lastly, a smaller number of studies have focused on teacher's favored perceptions of students rich in cultural capital (Dumais, 2006; Calarco, 2011; Malik, 2014). Yet Jæger and Breen (2016, pp. 1097) argue that very few studies have empirically tested the entirety of Bordieu's theory on cultural reproduction and in response put forth a dynamic model which includes both active parental investments and passive exposure of cultural capital in the childhood home. Results from this comprehensive model confirm findings from piecemeal studies that came before it. Namely, parents invest cultural capital onto their children which translates to their childrens' academic success, resulting in a feedback loop of parents' further investment (pp. 1108).

Micro-level qualitative analysis also provides compelling evidence for cultural reproduction. In her recent ethnography, contemporary theorist Jessica Calarco (2014) argues that parents of different classes instill distinct values in their primary-school aged children that lead to learned behaviors in the classroom to which teachers may be more or less receptive. Namely, children of working class parents are taught deference to authority and trust in decisions made by teachers and schools, whereas children of middle class families are coached values of entitlement—adopting an 'any-means' approach to problem solving (pp. 1016). As a result, working class pupils may be less likely to seek out help from teachers when they are struggling (Calarco, 2011). Inferred from this smaller-scale, ethnographic work is that cultural reproduction is sometimes explicit, and how well equipped a child is to satisfy culturally biased educational standards is highly dependent on class background.

The question that arises in sibship size research is: how does family size mitigate the processes of social and cultural reproduction? As summarized above, family size was once considered a robust predictor of level of education, with a higher quantity of siblings significantly reducing years of education (Blake, 1985). The issue with this finding is that family size is an endogenous variable and the degree to which the coefficient is correlated with the error term is unknown. Educated parents who wish to foster an ideal learning environment for their offspring may choose to create smaller families (Guo and Van Wey, 1999). In contrast, less educated parents may create larger families for a myriad of reasons: lower opportunity cost, restricted access or knowledge of birth control, orientation towards family values (Kreyenfeld and Konietzka, 2008). These key differences between parents of different family sizes might play a *more* integral role in shaping their child's preparation for obtaining higher educational attainment than that of family size.

This master's thesis poses the question: what if in isolation, family size *still* plays an integral role in social and cultural reproduction? A key element in explaining the relationship between sibship size and educational attainment is the allocation of precious parental resources. Resource dilution theory proposes that the birth of an additional child means a strain on parental attention, time, and money. If family size significantly predicts educational attainment in isolation, then larger sibships may obstruct the inheritance of fiscal, social, and cultural capital. Previous studies in RDT have focused on how increasing family size may contribute to the dilution of wealth inheritance (Lersch, 2019) and monetary resources imperative in the investment of higher education (Powell and Steelman, 1989). Yet, the dilution of cultural capital has seldom been considered in sibship size literature. Replicating a relatively novel approach first applied by Radyakin (2007) of using number of parental siblings and parental age of first birth as instruments to correct for family size endogeneity, this thesis hopes to isolate the effect of family size on both social and cultural reproduction. Thus, in addition to testing the direct effect of family size on educational attainment, the thesis regresses family size on the allocation of cultural capital itself.

Interpreting resource dilution theory through the lens of inheritance of cultural capital indeed seems to be uncharted territory in sibship size literature. This may be because neither cultural capital nor the specific resources in RDT are well-defined by their respective theorists. While Bourdieu (1976) references "statistics of theater, concert, and above all museum attendance" (pp. 57), he also distinguishes cultural capital into the broader categories of the *institutionalized state,* the *objective state,* and the *embodied state* (Bourdieu, 1986). The *institutionalized state* refers to educational qualifications, the outcome variable of this thesis, the obtainment of which requires both inherited and accumulated cultural capital in its two latter forms defined hereafter. These are: the *objective state*—the literal transmission of material objects such as paintings, books, or in modern context access to digital technology and the internet, and the *embodied state*—the "external wealth converted into the integral part of a person" through familial socialization (pp.18).

Parental resources are exemplified as time, attention, and material resources (Blake, 1989, pp. 11) but are rarely actually operationalized and directly regressed against sibship size in more recent literature. One exception is Blake (1989) who demonstrates evidence for sibship size having a negative correlation with participation in cultural activities such as dance lessons, music, and photography. Beyond this seminal work, however, few have investigated the impact of sibship size on cultural transmission directly. Finding evidence for both sibship size effects and the dilution of parental resources contributes an added credence to both the objectified and embodied state of cultural capital as a process of cultural reproduction. If cultural reproduction is not only dependent on passive exposure to cultural experiences, but also active investment of a child's cultural apparatus, then parents of multiple children might be burdened with difficult choices with limited means. While books may be shared, and music and sports can be enjoyed by multiple children at once, parents with multiple children (especially those with limited resources to begin with) may lack the money for golf and piano lessons for every child, or the ticket costs of multiple children to museums and plays.

Zajonc and Markus' (1975) confluence theory, which posits that families with closely spaced young children create an intellectually immature environment can also be easily integrated into the framework of cultural reproduction, particularly when considering the pathway of *passive* exposure of cultural transmission. This theory's focus on a child's intellectual environment based on the average, albeit dynamic intellectual milieu of each family member makes it possible to speculate its relevance to Bordieu's conception of inherited habitus. According to Bourdieu, a child is socialized into a habitus formed by the educational and cultural background of their caretakers. Yet, even if they are raised by parents rich in cultural capital, confluence theory predicts that the presence of too many children will inevitably create an environment less conducive to mature learning, and thus possibly less receptive to cultural transmission. As Zajonc and Markus (1975) argue, "some portion of intellectual growth will be determined by an interaction with the intellectual levels of their parents and their siblings" (pp. 76). Activities more sophisticated such as attending the theater or reading to a child may be replaced with more child-friendly, but less intellectually stimulating games and play with equally undeveloped children. There are many valid criticisms of confluence theory which will be detailed in chapter two. Barring these criticisms, however, even without consideration of cultural capital in the form of the objectified state, family size may have theoretical implications for the embodied state and cultural reproduction if evidence is found in support of confluence theory.

As addressed previously and further detailed chapter two, family size research has been criticized for neglecting selection bias and the endogeneity of its primary explanatory variable. However, with adjustments for selection bias pioneered by Radyakin (2007) and Jæger (2006), this thesis argues that family size research warrants a re-evaluation due to its contributions to a broader theory of social reproduction. There is no doubt that socio-economic status, race, sex, and immigration background play an integral role in social reproduction. Yet, with consideration to evidence presented by Radyakin (2007), there is reason to believe that family structure might play a mitigating role on these influences.

1.2.2 The German Context

Understanding the German context of educational inequality and the role of family structure may be aided in its placement in Esping-Andersen's (1990) welfare regime typology. In his seminal work *The three worlds of welfare capitalism,* Esping-Andersen (1990) categorizes nations into three typologies of welfare states: the *liberal regimes*, *social-democratic regimes*, and the *conservative regimes*. The liberal regimes, consisting of most anglo-saxon countries such as the United States and United Kingdom, are defined by low degree of decommodification, low-levels of state expenditure, and an emphasis on the free market solutions and meritocracy as an alternative to welfare redistribution. The social-democratic regimes, which include most Nordic Scandinavian countries, are conversely defined by high-levels of decommodification, high-levels of state expenditure, and universal redistribution of welfare. Germany and the remainder continental Europe fall under conservative regimes. Sometimes referred to as corporatist regimes, conservative regimes are characterized by their ideological devotion to maintaining rigid social structures, moderate state expenditure compared to the nordic model, and redistribution limited to "recipient's record of contribution" (Peter et al., pp. 246). Before moving forward, it is important to disclose that reliance on Esping-Andersen's framework comes with its limitations. Esping-Andersen (1999) himself writes that use of typology comes "at the expense of nuance" since these models are "especially static," yet when taken cautiously they aid the analyst to "see the forest rather than myriad trees" (pp. 72).

Several studies attempt to extrapolate welfare state typologies into understanding its implications on educational policy. West and Nikolai (2013) exemplify Germany as a continental educational regime with high degree of stratification, above average differences between high and low reading scores, and relatively low public expenditure on education compared to nordic countries (pp. 482). In regard to tertiary education, Pechar and Andres (2011) identify conservative welfare regimes as among the lowest in higher education expansion based on enrollment rates as well as lowest in public expenditure despite no tuition fees. This finding is echoed by Dodin et al. (2024) who find that educational expansion (*Bildungsexpansion*) in Germany has only led to uniform increases in A-level educated pupils across the income gradient, leaving inter-class inequality unaffected (pp. 2). In conducting a cross-national analysis of the degree of educational inequality grouped into welfare regimes, Peter et al. (2010) find that conservative regimes rank highest in between-school educational inequality even above free-market liberal regimes.

There are two interrelated explanations for these findings. The first is Germany's secondary school structure. While German pupils are pooled together in primary school (*Grundschule)* from grades one to four, in grade five children are sorted based on academic ability into three different types of schools: lower secondary school (*Hauptschule*), intermediate secondary school (*Realschule*) or upper secondary school (*Gymnasium*). Each of these schools vary in length, rigor, and opportunity structure. As summarized by Pietsch and Stubbe (2007), *Hauptschule,* with a duration of five years, leads to the attendance of vocational school (*Berufsschule*), whereas *Realschule* leads to six years of education and offers the additional opportunity to attend a more highly-skilled technical school (*Fachschule*). The upper secondary school (*Gymnasium*), which takes eight or nine years to complete (dependent on the Federal State), leads to the obtainment of a diploma (*Abitur*) and the possibility to attend an academic institution such as a *Universität* or *Hochschule*. Although rare cases of inter-track permeability have become possible in different federal states (Schneider, 2008), theory shows that the existence of early selective sorting in conservative regimes is strongly responsible for replicating social inequalities (Jonsson and Erikson, 2000; West and Nikolai, 2013). This is reflected in Pietsch and Stubbe's (2007) finding that the probability of children of semi-skilled and manual workers attending *gymnasium* are 10% compared to the chances of children of higher professionals at 70% (pp. 439). Since secondary school sorting is often determined by teacher recommendation, Germany's system of segregation provides interesting context for testing Bourdieu's theory of social reproduction discussed in the previous section.

The second possible explanation for in-between school inequality is that education policy is sanctioned to Germany's sixteen individual federal states (*Bundesländer).* To this end, permeability between school tracks, parental influence in secondary school sorting, age of first sorting, and duration of gymnasium are all ways in which educational policy varies between federal states (Klein et al., 2008). Despite findings of regional variation of social mobility between federal states, Dodin et al. (2024) argue that respective state differences in secondary school tracking policy do not readily explain their findings, but rather Germany's historical east-west divide since reunification (pp. 12). On the other hand, in tracking the state of inequality of opportunity (IEO) surrounding the reunification process, Klein et al. (2008) find that levels of IEO significantly increased in the former GDR after reunification, converging with levels of West Germany. They argue that these institutional changes benefited children of *Abitur*-holding parents, especially in former east German states such as Thuringia and Saxony, where teachers had binding influence over parents on a child's tracking decision (pp. 18).

Determining whether regional differences in school-tracking policies significantly impact regional inequalities is beyond the scope of this thesis. Nevertheless, the presence of early school tracking and regional variation in educational inequality make Germany a very interesting case-study for empirically understanding how educational policy may impact cultural and social reproduction. On the one hand, if a pupil lives in a federal state in which parents may overrule teacher recommendation on school sorting, then the teacher's perception of a pupils cultural capital may have less salience. Working class parents may have more agency to help their children achieve beyond their own achievements. On the other hand, as evidenced by Calarco

(2014), working-class parents may have internalized deference to authority, taking teacher's recommendation at face value, rather than challenging their potential bias. In either regard, despite providing its citizens state-funded, tuition-free higher education, access to that education is heavily gate-kept based on class-background (Pietsch and Stubbe, 2007) and migrant status (Becker and Reimer, 2010). Financial capital is not the defining mechanism that filters out students from low-income families in pursuing an academic qualification such as in liberal regimes like the United States, but rather the institutional-design of educational institutions itself which favors families already rich in cultural capital. In other words, compared to most other western nations, cultural capital may play a larger role in educational outcomes than capital itself.

Beyond educational policy, Esping-Andersen's welfare typology also provides insight to the second relevant factor of this thesis: family structure and fertility rates. To describe the influence of the welfare state on family formation, Esping-Andersen (1999) refers to each welfare regime's level of *de-familialization*—referring to the degree to which the state relieves the financial and time-related burdens of raising children from the individual household (pp. 51). Policies of de-familialization include increased public child-care, public spending on family services, and tax-subsidies of child families—a list that is descriptive but by no means exhaustive (pp. 61). Esping-Andersen finds bimodal results when examining de-familization by welfare state: with Scandinavian countries occupying one extreme, and the remaining regimes occupying the other among which results only differ "by an accent" (pp. 62). Esping-Anderson (1999) also reports a paradox in which the most familistic regimes (e.g. Italy and Spain) host some of the world's lowest fertility rates, while Scandinavian countries, with universal generous welfare provisions that benefit working mothers rank highest in fertility among Europe (pp. 67).

From his conception of de-familialization, it is evident that the policies that influence family structure are centered around the rational choices made by women, in a society in which women are outpacing men in human capital accumulation. Therefore, the degree of de-familialization in Germany is relevant to our thesis in two regards. First, policies of de-familization may impact the degree to which limited resources which are diluted among multiple children may be outsourced by the state. If resource dilution theory relies on the assumption that childcare responsibilities lie solely on the family (or more specifically the mother), then a welfare state that boasts a strong emphasis on de-famlialization may mitigate this dilution of resources, rendering sibship size effects less relevant. Second, the degree to which the state deburdens the mother's childbearing responsibilities may affect the opportunity costs of an educated, employed mother and thus further impacting her family planning decisions. This is particularly impactful on selection bias proposed by Guo and Van Wey (1999), which argues that highly educated mothers may choose to create smaller families.

Germany's degree of de-familialization and fertility rates by level of education must therefore be considered in the context of my analysis. Kreyenfeld and Konietzka (2008) find that while fertility rates overall have decreased in Germany since the 1970s, a closer look finds a closing gap in fertility rates between higher- and less educated women—in particular, highly educated women are marginally having more children on average, and less educated women are having fewer (pp. 174). One explanation given for this is a "time-squeeze effect" which proposes that women who obtain a higher education put off starting a family, but then in fighting against their biological clock have more children in quick succession after obtaining an education. Evidence for this is inconclusive, with Kreyenfeld (2002) finding stronger evidence for two alternative theories: first, that highly-educated women are more likely to select highly educated partners with greater earning potential for larger families; second, that family-oriented graduates select themselves into groups at higher risk of second and third births. Regardless of the explanation, these findings contradict the selection bias proposed by Guo and Van Wey (1999).

Falling fertility rates, especially among less educated mothers may be explained by Germany's middling levels of de-familization in Esping-Andersen's (1999) analysis. Yet in tracing the nation's history of policy reform, Ostner (2010) finds that after a "West German male-breadwinner model" overtook an "East German dual-worker model" during reunification, Germany has shifted gears since 2002 to a "sustainable family model" as an effort to reverse fertility rates towards Scandinavian levels (pp. 211). Policies under this ideological shift replaced cash transfers to young parents with an expansion in public services such as extended free daycare to incentivize mother's to stay in the workforce. In addition to offsetting the opportunity cost of working women having children, the aim of these policies were to address concerns of increasing “child poverty”—that is “scarcity of parents resources…money, space, time, opportunities”as well as “qualitative shortcomings such as stimulating environments” the private household supposedly fails to provide (Ostner, 2010, pp. 226). This raises the question, what implications do these policy shifts have on sibship size effects? As previously discussed, an expansion of public daycare and an emphasis on supporting working mothers strongly correlates with theories proposed in this thesis. If placing children in public daycare is intended to supplement resources low-income parents cannot provide, how might this mitigate cultural reproduction? If daycare settings are meant to create “stimulating environments” (pp. 226) private households might not provide, where does this fit into Zajonc’s confluence model? Lastly, if highly-educated working mother’s need not make great sacrifices to have children, how might selection of family size change?

As outlined in this section, Germany provides a particularly pertinent political landscape to test the effects of family structure on educational attainment. On the one hand, as typified as a conservative welfare regime, Germany’s early sorting and strict stratification makes it ripe for educational inequality by design. This may be unexpected when comparisons are drawn against heavily commodified, high-cost tuition universities in liberal regimes; but, despite the wealth of state-funded, tuition-free academic institutions, access to such heights tend to fall along culturally biased, class lines. On the other hand, Germany’s welfare state has slowly shifted toward a nordic model in terms of family policy, aiding the compatibility of motherhood and full-time employment. While fertility rates are still low, such policies should have profound implications on family structure and pooling resources that may otherwise be scarce in larger families. If sibship size effects are found to contribute to cultural reproduction when controlling for selection, these contradictory policies regarding family and education may provide sound explanation.

1.3 State of the art

1.3.1 The problem with sibship size

Studying the effects of family size and sibling composition is not a novel endeavor. In their literature review, Powell and Steelman (2001) argue that “sibling configuration is as old as sociology itself.” Modern efforts in understanding the effects of family size originate in Blake’s

(1989) seminal work in which she examined the effect of sibship size on educational outcomes in multiple large cross-sectional dataset in the US. Since this publication, sibship size has been found to be influential on grades (Powell and Steelman, 1990), probability of graduating (Jacob, 2011; Paar, 2006: Kalmijn and van de Werfhorst, 2016), intellectual ability (Ghilagaber and Wänström, 2015; Jæger, 2006), and school attendance (Feng, 2021).

While findings for sibship size effects have been replicable and persistent in large cross-sectional, critics in the late 1990s brought into question the heavy-handed causal interpretations of their contemporaries. Specifically, Guo and Van Wey (1999a) tested the replicability of these findings using two dynamic approaches. First, they tested sibship sizes using change models, in which changes in cognitive ability test results were examined between periods in which the respondent's household was introduced to a younger sibling. Second, he applied sibling fixed effects in which cognitive test results of the eldest child were compared to those of the youngest child. Both techniques revealed that cross-sectional designs heavily conflated family size estimates. A year later, Rogers and Cleveland (2000) applied similar within-family analysis, yielding similar insignificant results regarding birth order effects. The conclusions made by these scholars was that family size is an endogenous variable and that cross-sectional designs underestimate unobserved time-variant differences in fertility decisions and family planning preferences between parents even net of multiple control variables. In short, these scholars determine that small families do not so much make intelligent families as intelligent parents make small families.

1.3.2 Replicating Radyakin (2007)

Following the suggestion by Jæger (2006), Radyakin (2007) addresses the problem of family size endogeneity by introducing four instruments into his analysis: mother's number of siblings, father's number of siblings, mother's age at first birth, and father's age at first birth. The methodological strategy to using these instruments is detailed in section three of chapter two, but in summary, when two people meet the reproductive capabilities of their partner (indicated by number of siblings and age) is unknown to the other and therefore qualifies as a random event and therefore exogenous to the outcome variable (Jæger, 2006). Furthermore, the amount of siblings one has may culturally imprint on each spouse the number of children they choose to have. Therefore, use of these instruments should attenuate family size endogeneity and provide relatively unbiased results. Radyakin's (2007) results not only validate the significance of sibship size effects, but boost the magnitude of the effect by a critical margin.

Amidst a replication crisis in the field psychology, Freese and Peterson (2017) draw attention to the fact that sociology is left out of the conversation. The intended contribution of this thesis is not only to replicate Radyakin's (2007) seventeen year-old findings and consequently bring social science into the conversation. Instead, the intention of this thesis is to expand on his findings and address their shortcomings. This is achieved in three areas. First, while the same dataset will be used in analysis, there has been a significant increase in data availability and sample size. This alone allows a robustness check to Radyakin's (2007) relatively small sample size. Second, I extend the scope of analysis by regressing adjusted sibship size on relatively newly implemented results from a cognitive ability test as a means of testing confluence theory. Lastly, I use Radyakin's (2007) instrument to test various measures of cultural capital to validate the mechanisms proposed in both resource dilution theory and cultural reproduction.

1.4 Thesis Outline

Sibship size research, that is the effects of family size on educational attainment, has fallen out of vogue in the field of social structure. While the 1990s saw a genesis and proliferation of cross-sectional evidence for a tradeoff between child quantity and child quality, skeptics drew attention to the endogeneity of the primary explanatory variable and its resulting selection bias leading to a shift away from inter-familial inequality research (Guo and Van Wey, 1999: Rodgers et al., 2000). The conclusion made from this criticism was that the ultimate driver of inter-familial educational inequality was not family size, but factors related to socioeconomic status itself. The aim of this introductory chapter was to place sibship size literature within a grander context of cultural reproduction and the complimentary German political landscape in order to justify its re-evaluation. The conclusion made from this chapter is that the increasingly popular treatment for independent variable endogeneity, instrumental variable analysis, provides a warranted tool in re-examining family size effects within the theoretical framework of social reproduction. The remainder of this master's thesis is structured as follows. Chapter two will detail the three driving theories of sibship size effects and how they inform our hypotheses. A particular focus will be made on the selection bias theory, to explain how the instrumental variable for this analysis was chosen. Chapter three will introduce the data and variables chosen to replicate Radyakin's (2007) findings. Additionally, this chapter will explain the methodology replicated from Radyakin (2007) and the slight deviations from his method of analysis. Chapter four will present the results from our analysis, including expansions made from Radyakin's (2007) work. Lastly, chapter five will discuss these findings as well as their shortcomings that may be corrected for in future research.

Chapter two: Theoretical Framework and Hypotheses

In the previous chapter, the two mainstream theoretical explanations for family size effects were briefly introduced. This chapter will elaborate in greater detail their mechanisms, caveats, and limitations in order to form coherent hypotheses used to answer the driving questions of this thesis. Additionally, the last section will provide greater context to the overriding selection problem with family size research and introduce the various instruments proposed to alleviate it. This will include the justification of my chosen instrument, and the final set of hypotheses which it informs.

2.1 Resource Dilution Theory

As introduced in the previous chapter, one theory used to explain the negative correlation between number of siblings and educational outcomes is resource dilution theory (RDT). First developed by Judith Blake (1981), she observes that the growth of the family means the greater division of resources provided by the parents. As a primary aim of the thesis is to understand the mechanisms of the child quality-quantity trade-off it is important to address two questions regarding RDT: (1) Which resources do parents provide to their children? (2) How might different resources be allocated? In addition to exploring these questions, section 2.1.2 will address the various challenges propped against RDT.

2.1.1 Mechanisms

According to Blake (1981), parents provide their children resources in the form of three broad categories: *environments, opportunities*, and *treatments* (pp. 422). *Environments* or *settings* take form in the "types of homes" and "necessities of life" but also, as addressed in the introduction, "cultural objects" provided by parents (pp. 422). Blake (1989) clarifies that parents foster their child's environment by providing them enough individual privacy to learn and develop as well as by granting adequate access to a collection of reading material—or contended through a modern lens, access to reliable internet (pp. 11). As will be discussed later in this chapter, environment also plays a prominent role in Zajonc's confluence model, where an environment conducive to productive development is directly dependent on family structure. In the context of resource dilution theory, however, environment is informed and operationalized most directly by material and cultural resources rather than sibling configuration itself. That is, there is a proposed indirect effect on child quality outcomes, where income and wealth are diluted with each additional child which impedes the parents' ability to provide the aforementioned privacy and cultural objects. While there has been substantial evidence for the dilution of economic resources such as wealth (Lersch, 2013) and financial contributions for university (Powell and Steelman, 1989), there has been little to no observation of the dilution of resources more particular to the environment such as educational materials notwithstanding some notable exceptions (Downey, 1995; Teachman, 1987).

A similar problem plagues *opportunities*—defined by Blake (1981) as the chances for children to acquire knowledge and experience from the external world. In the context of cultural transmission, these are the *embodied state* of Bourdieu's theory on cultural capital—the opportunity to travel, attend concerts or theater, or to engage in expensive sports. There is definite theoretical overlap between opportunities and environment; parental socioeconomic status and wealth are antecedents to both types of resources, but while environment focuses on resources *inside* the home (namely educational materials, technology, and the home itself) that aid learning and intellectual development, opportunities focus on the resources parents provide their children *outside* the home—an accumulation of cultural capital. Just as with measures of environment, there is little empirical testing of the dilution of opportunities, especially in recent years. Among the few exceptions, scholars operationalized opportunities as participation in cultural activities (Blake, 1989; Downey, 1995).

Parental treatment refers to the social resources provided to a child—time, individual attention, encouragement, and parental involvement (Blake, 1981). Similar to resources that shape setting and opportunities, it could be argued that wealth and income in and of themselves could supplement these resources. Parents from wealthier families may have more financial leverage to take maternal and paternal leave in their child's younger years in order to provide their child with individual attention and time. Additionally, wealthier parents could outsource allocation of social resources to child caretakers and private tutors. Yet high socioeconomic status compensating for poor parental treatment is no guarantee, as parents affording time with their children is not always followed by encouragement, and hiring of caretakers does not provide children with parental consistency. It is therefore argued that allocation of social resources should also be measured discreetly from that of material resources. Blake (1989) finds in her seminal work that a child's perception of parental involvement and encouragement is more influential on educational outcomes than resources associated with setting (pp. 439).

Table 2.1 Types of resources in RDT and their indicators.

Resources	**Measures**
Environment	Income, Wealth, Privacy, Educational materials
Opportunities	Income, Wealth, Cultural activities
Treatment	Time, Attention, Involvement

While there are many empirical reports of the correlation between sibship size and educational attainment, much more unexplored is the direct dilution of specific resources with the growth of a family and its mediating role on educational outcomes. Dilution of material resources is most frequently found in literature. Lersch (2019), for instance, finds that the birth of an additional sibling reduces wealth by 38 percent. Similarly, Powell and Steelman (1989) find that in the US context, having additional brothers reduces parental financial contribution to higher education by 27 percent (pp. 140). These findings, however, do not fully capture the mechanisms of *environment* and *opportunities* since access to material wealth is only a proxy for a child's acquisition of educational materials and participation in cultural activities. As mentioned in the previous chapter, Blake (1989) measures the dilution of opportunity as the participation of cultural activities such as dance lessons, music, and photography.

Downey (1995) provides the most holistic approach in understanding the distinct mechanisms of RDT by measuring the dilution of *economic resources*—most akin to opportunities and environment i.e. computer in the home, educational objects, money saved for college, cultural objects, and cultural activities—as well as *interpersonal resources*—measuring treatment through frequency of talk, familiarity with child's friends, and familiarity with parents of child's friends (pp. 751). Strohschein et al. (2008) expand on Downey's findings by contributing a dynamic approach to understanding the dilution of resources. Specifically, they focus on the dilution of social resources by measuring the presence of maternal depression and family dysfunction at different cross sections as families increase in size.

Results from Downey (1995) and Strohschein et al. (2008) also suggest that the allocation and dilution of resources are more complex than simple linear division. Downey (1995), for instance finds that economic resources such as money saved for university, cultural classes, and cultural activities decrease much more rapidly than other resources—following a *1/x* function with *x* representing the number of children in a family; this contrasts starkly to social resources which follow a simple linear function (pp. 755). Strohschein et al. (2008), on the other hand, yields findings consistent with RDT with the reduction of positive interactions,yet also find no evidence for the dilution of consistent parenting. Their explanation for this is that social resources are reallocated rather than diluted to meet the needs of every family member with each new additional member.

From these two studies, it is evident that RDT is a more complex process than proposed by Blake (1989). While there is clear cross-sectional evidence binding family size with educational outcomes, little has been done to investigate the process of dilution. For this reason, RDT has come under considerable criticism.

2.1.2 Caveats and Criticism of RDT

Several assumptions of resource dilution theory are central to its criticism. The first assumption is that resources are allocated uniformly among children. Challenges to this assumption come from two arenas of literature. First, a number of studies have found that allocation of resources is based on gendered preferential treatment (Powell and Steelman, 1989; Powell and Steelman, 1990; Chu et al., 2008; Jacob, 2011; Kalmijn and van de Werfhorst, 2016). Powell and Steelman (1989) conceptualized this as the "liability of having brothers" whereby competition between siblings for resources is steeper for individuals with more brothers regardless of the focal person's gender.

The explanation follows that in a patriarchal society, following rational choice theory, parent's human capital investment in sons will provide higher returns than if they were to invest in the education of their daughters. Evidence for this theory is supported by Kalmijn and van de Werfhorst (2016), who find that the "liability of having brothers" is stronger in countries which rank higher in the gender inequality index. In the German context, Jacob (2011) finds that birth order plays an intersectional role in gendered sibship size effects; while having a brother has no significant effect generally, having an *older* brother has a significant negative effect on a woman's chances in graduating from a university or *Fachhochschule*. Conversely, Chu et al. (2007) finds that in Taiwan it is especially detrimental for a woman to have younger brothers, speculating that women are sent to work in order to provide capital for their younger brothers to attend university. While evidence for gender asymmetry and the liability of having brothers is nuanced and heavily dependent on cultural context, it suggests that resource dilution does not impact every child in the family equally.

Additional evidence for unequal allocation is found in sibling correlations literature. The intent of studying sibling correlations is to measure inter-familial differences in educational attainment in order to approximate how much educational inequality can be attributed to family of origin effects. The underlying assumption in sibling correlations, however, is not only that educational outcomes differ somewhat between siblings, but also that parents strategically invest more in children with weaker chances at educational attainment—sometimes referred to as compensatory class hypothesis. The argument of this theory follows that families from higher class origins are better able to compensate for ability differences between their children (Grätz, 2015). Evidence for class compensatory hypothesis is nevertheless inconclusive. On the one hand, Grätz (2015) finds that wealthier families are able to compensate for the adverse effects of parental separation on educational outcomes such as attending *gymnasium* and obtaining adequate marks. On the other hand, there have been very little significant differences in sibling correlation coefficients when stratified by families of high and low socioeconomic status (Grätz, 2018; Duta and Breen, 2021). Despite this, compelling evidence suggests that birth order effects, that is the conjecture that higher birth siblings are advantaged over later born siblings, is smaller in higher class families (Grätz, 2018). Whether the compensatory class hypothesis is a reliable theory is not a question this thesis aims to answer. However, it provides relevant insight into how resource dilution should be considered.

A second assumption of the resource dilution hypothesis challenged by critics is that resources are perfectly finite. As already demonstrated in the previous section, evidence suggests that different types of resources are allocated at different rates (Downey, 1995). In addition to this, some literature suggests that resources are also able to be pooled, regenerated, and redistributed (Strohschein et al, 2008; Sun and Li, 2008; Shavit and Pierce, 1991). An example of the latter function is presented by Sun and Li (2008), who find that larger sibships attenuate the negative effects of divorce on test scores, inferring that older siblings may supplement for emotional and attention resources when parents are unavailable. This finding is echoed in the previously mentioned explanation for Chu et al.'s (2007) conclusion, that older sister's generate additional financial resources for their younger brother's to obtain higher education. Shavit and Pierce (1991) contribute to this proposal by drawing attention to cultures in which the responsibility of child-rearing extends beyond the nuclear family, thus under certain conditions, attention and time are perhaps not as perfectly finite as theory predicts. The overarching argument found in these studies is that resource dilution underestimates the importance of interfamilial (particularly sibling) interactions. However, the effects of interfamilial interaction and older siblings acting as co-parents is based on conjecture in the aforementioned studies rather than evidence from direct testing. Furthermore, failings of RDT to account for interaction between family members is compensated for in confluence theory, which is discussed in the following section.

Another debated assumption is that age spacing plays a significant role in resource dilution. While traditionally considered a mechanism of confluence theory, Powell and Steelmann (1990) argue that children benefit from wider age spacing (or larger sibling density) because parents have more time to recover lost resources between children. Accordingly, close spacing is apparently especially detrimental to financial assets but can also impact social resources such as time and individual attention even when accounting for economies of scale (Powell and Steelman, 1990 pp. 185). However, Jæger (2009) points out that stronger evidence supports age spacing as a factor predicting cognitive ability under the mechanisms of confluence theory rather than educational outcomes, and that the former acts as a mediator of the latter in resource dilution. Similarly, while some theorists argue that first born children have an advantage in resource dilution since they have their parent's undivided attention for some time before the birth of their younger children, in a review of the development of sibship size literature, Steelman et al. (2002) find birth order effects "no consistent pattern…on intellectual performance or educational attainment" (pp. 257). Whether this is true, Jæger (2009) makes a sound theoretical argument of why neither mechanisms seem to have relevance in RDT. He posits that while resource dilution may impact cognitive ability, it should also have an "additional effect of sibship size on educational attainment" unrelated to intellectual performance (pp. 4). In other words, while social resources such as parental attention may impact cognitive ability which mediates educational attainment, fiscal resources should have a direct effect on educational attainment independent of intellectual milieu. As acknowledged in the introduction, Germany is a country with tuition-free university education and is dominated by public schooling. However, well established is that family background and socioeconomic status play a significant role in educational outcomes despite this illusion of barrier free provisions. For this reason, age spacing and birth order effects are excluded from empirical tests on RDT.

2.2 Confluence Theory

If resource dilution theory underestimates the role of interfamilial socialization, confluence theory provides a model that distinctly accounts for this feature. Developed by Zajonc and Markus (1975), confluence theory takes a social psychological approach to understanding sibship size effects. Accordingly, it determines the average cognitive ability of each person in the family and how this influences the degree of intellectual stimulation of a family setting—predicting that larger families, particularly ones with small children, create an environment less intellectually stimulating for each family member leading to lower levels of mature intelligence. Not accounted for by confluence theory is educational attainment. In developing a joint-model to harmonize both theories, Jæger (2012) argued that while confluence theory only determines cognitive ability, resource dilution theory accounts for a broader spectrum of child outcomes including intelligence. In other words, intelligence can be seen as an outcome variable for confluence theory as well as a mediator for educational attunement under resource dilution models. Therefore, a valid test of confluence theory should consider the mechanisms of CT and their relationship with intelligence. The following sections will discuss the relevant mechanisms of CT and address its criticisms.

2.2.1 Mechanisms

Key in understanding the mechanisms of confluence theory is comparing its view of family setting compared to resource dilution theory. As mentioned in a previous section, RDT sees family environment as a product of parental contribution to educational materials and cultural objects. Confluence theory, on the other hand, views intellectual environment as the household's average intellectual level determined by each member and unadjusted for age (see equation 1 below). In this way, other members of the family serve as dynamic resources to each other in cognitive development and their interaction is the mechanism informing this relationship. This is found in empirical testing considering measures of stimulating intellectual activities such as childhood reading, time spent alone, and time spent with friends—all predictors of verbal intelligence (Mercy and Steelman, 1982). While all interaction is essential to verbal and language development, confluence theory posits that children benefit the most from communication with fully developed adults rather than interaction with other children (Mercy and Steelman, 1982, pp. 540). Therefore, the ratio of adults to children in the household is important for cognitive development.

$$Intellectual\ Environment\ = \frac{\Sigma\ Intellectual\ Levels\ of\ Family\ Members}{Number\ of\ Family\ Members} \qquad (1)$$

One implication of this is the importance of birth order. First-borns and children without younger siblings have the greatest advantage according to CT since the intellectual environments in which they are (in the former case, at least partially) raised is dominated by cognitively developed adults. Zajonc and Markus (1975) argue that the addition of a new baby dampens the intellectual level of every member of the family, including that of the parents—citing the regression of kindergarten teacher's verbal ability (pp. 84). Thus, as the family grows, younger children begin to dominate the intellectual environment replacing intellectually stimulating activities such as reading and time spent with adults with time spent watching television or with peers in similar phases in development (Mercy and Steelman, 1982). To explain why first-born children sometimes tend to perform better academically than children without siblings, Zajonc and Markus (1975) introduce the teaching mechanism, which proposes that older siblings benefit from tutoring their younger siblings. An example of this would be an older sibling helping their younger sibling with a homework assignment they had completed themselves a year or two before. While this benefits both parties in the interaction, this practice especially helps the first born in reinforcing knowledge he already attained (Falbo 1978, pp. 322). Therefore, in congruence to family size effects, confluence theory argues that birth order plays an essential role in explaining cognitive ability.

A second implication of confluence theory is that the age spacing between siblings is important in shaping a family's intellectual environment. The first-born advantage of monopolizing their parent's time and attention as discussed above is dependent on the amount of years until the birth of the next child—with closer spacing strengthening the teaching advantage. Moreover, the necessary dilution of a household's intellectual environment that follows the birth of an additional child is possibly mitigated with larger age spacing between children (Zajonc and Markus 1975, pp. 83). This is best explained with simple algebra functions.

$$Family\ A = \frac{30 + 30 + 5 + 0}{4} = 16.25 \qquad (2)$$

$$Family\ B = \frac{30 + 30 + 25 + 0}{4} = 21.25 \qquad (3)$$

$$Family\ C = \frac{30 + 30 + 0}{3} = 20 \qquad (4)$$

Derived from the first equation (1) are three three hypothetical families. *Family A* and *Family B* represent the same family size with one key difference between them: age spacing. Since *Family*

A is spaced closely together, with the developing first-born toddler having an intellectual level of 5 and the infant having an intellectual level of 0, the resulting intellectual milieu is only 16.25. Inferred from this hypothetical is that the last-born benefits from being in a widely spaced family (*Family B)* because they are raised in an intellectual environment five points stronger than a closely spaced family (*Family A*). More surprisingly, the youngest child of a widely spaced family benefits more from their environment, than one from a single-child family *(Family C*) despite them both suffering from a "teaching handicap"—having no younger sibling to mentor (Zajonc and Markus, 1975, pp. 83).

A third implication of confluence theory that is often unexplored in the literature is that the number of adults in the household should counteract family size effects. Thus as suggested by Shavit and Pierce (1991), multi-generational households and community parenting not only benefit children in terms of regeneration of resources (as implied under RDT), but also in enriching the intellectual environment. For this reason, I argue that there is significant theoretical overlap between CT and RDT. A common idea is that CT acts as a subsample to RDH since the intellectual environment places an emphasis on the value of interactions between cognitively mature adults and developing children. This could be reinterpreted under RDH as the abundance or lack of interpersonal resources (i.e. treatment) of the parents. However, Jæger (2009) challenges this assumption by arguing that the mechanisms of age spacing, birth order, and sibling teaching as an exclusive influence on intelligence makes CT theoretically distinct from RDT. Therefore, in line with Jæger's joint test, we apply these three mechanisms as a separate test of confluence theory to explain variation in cognitive ability.

2.2.2 Caveats and criticism of CT

Despite initial enthusiasm for the confluence model, strong empirical evidence for the theory failed to present itself (Galbraith, 1982; Page and Grandon, 1979; Valendia, Grandon, and Page, 1978). One common thread from these studies is that while aggregate data generally produced results consistent with CT, effects of family size and birth order effects disappeared under individual level analysis (Valendia, Grandon, and Page, 1978; Page and Grandon, 1979). At best, family size effects lost statistical significance on an individual level analysis; at worst, reverse trends were observed from what CT predicts. An additional challenge made to CT was in regards to age spacing. Galbraith (1982) for example, finds that the age spacing required to generate an intellectual environment superior to that of his/her older sibling stretches beyond the child-bearing years of even the youngest of mothers. By his calculations, age spacing as a mediator of family size effects has very unrealistic theoretical grounds.

To defend the validity of their model, Zajonc (1983) asserted that the previous tests of the confluence theory applied linear regression on non-linear data and failed to account for the dynamic nature of a family's intellectual environment. Accordingly, conducting an appropriate test for confluence theory requires use of "iterative calculations based on… nonlinear least squared methods", the inclusion of parental intelligence level, and repeated measures of a family's intellectual environment "until the last child reaches maturity" (pp. 465). From this demand for increased standards, the testability of the confluence model has been brought into question. Steelman (1985) contends that Zajonc's model, with impractical demands for longitudinal data of the entire family "comes perilously close to becoming untestable"; that these amended standards fail even to match up to Zajonc's original research, and that confluence theory should be used as a springboard in understanding sibling configuration effects by way of developing new models (pp. 378). While the present thesis uses a longitudinal dataset that accounts for every member of the household, Zajonc's standards are still unable to be met as data on cognitive ability is neither measured of each family member nor on multiple occasions. To this extent, testing the dilution of the time-variant intellectual environment directly is not within the means of the present thesis. Instead, my use of linear regression with instrumental variables to test the mechanisms of CT (age spacing, birth order, and teaching mechanism) is justified in its effort to correct a graver, more fundamental threat to sibship size literature: selection bias.

2.3 Selection Bias and Family Size Endogeneity

2.3.1 The Problem with Causality

One of the most existential challenges to sibship size literature is the theory of selection bias. This theory proposes that family size effects are neither a result of resource dilution nor poor intellectual environment, but instead a methodological artifact resulting from selection bias. Problems with selection bias were detected when Guo and Van Wey (1999) and Rodgers et al. (2000) respectively attempted to test the causality of family size effects on intelligence.

A common thread among critics of family size literature is that too often cross-sectional models were used to assume causality. To challenge these assumptions, Guo and VanWey apply sibling fixed-effects models and change models to predict cognitive ability—finding that statistical significance disappears and, in some cases of significance, the coefficient of sibship size becomes positive. Their explanation for this is that "time-invariant family influences" are captured in change models that are not captured in conventional analysis (Guo and VanWey 1999, pp. 182). One such influence hypothesized by the authors is the parents' orientation toward cultivating a rich intellectual environment, suggesting that parents with this orientation may opt for having a smaller family (pp. 183). Similarly, Rodgers et al. (2000) after yielding similar results using change models, coin birth order effects as a "methodological illusion" attributed to a plethora of selection biases undetected in the use of cross-sectional analysis conventional to previous studies on this topic (pp. 599). In regards to family size, the authors conclude in agreement to Guo and VanWey (1999) that small families make not intelligent offspring, but that intelligent parents make small families. In other words, it is difficult to isolate the direct effects of sibling configuration on educational outcomes because the variable at hand is endogenous; linked to the aforementioned covariant influences such as family size preferences and career ambitions.

Pioneers of family size effects directly challenged these critiques in light of their respective publications. Downey, Steelman, Powell, and Pribesh (1999) (hereafter Downey et al.) assert that Guo and VanWey's (1999) rejection of sibship size effects was made prematurely on the basis of four distinct arguments: neglect of age spacing, implausibility of covariate coefficients, the unfair comparison of model structures, and an unpersuasive theoretical explanation. Guo and VanWey (1999a) convincingly thwart the first two arguments made by their challengers. First, as argued previously in this thesis, evidence for the effects of spacing, although theoretically relevant to the confluence model, are not entirely consistent. Net of this argument, however, Guo and VanWey (1999a) present evidence that distinguishing between effects of having siblings close and far in age is still washed out in both change models and sibling analysis. Second, Downey et al. (1999) allege that the statistical insignificance of time-varying control variables such as socioeconomic status seem implausible. In rebuttal, Guo and VanWey (1999a) maintain that since the focus of the study was family size, the change models were designed to maximize the variation of family size rather than the variation of time-variant control variables such as mother's education and therefore not informative to the effects of such control variables (pp. 203).

The first two critiques made by Downey et al. (1999) are reasonably and constructively combated by Guo and VanWey (1999a). Their latter two arguments, however, raise questions that the present thesis aims to answer. First, Downey et al. (1999) argue that there is too little variation of time-variant factors between the observed years of 1986 and 1992 to detect family size results—claiming this leads to unfair comparisons between cross-sectional models which compare families of different sizes while controlling for other family differences. The second argument made is that the theoretical explanation of parent's orientation toward learning is static and underestimates the dynamic process of family planning (pp. 196). While the present thesis does not deny the validity of Guo and VanWey's findings, specifically grounds for non-causality, it does question whether selection bias fully accounts for previous cross-sectional findings. In line with Downey et al. (1999), this thesis proposes a more congruent model to cross-sectional data not to prove causality, but instead to test whether dilution effects persist net of family size endogeneity. In the following section, we will detail theoretical explanations for different variations of this model, and how the present thesis selected the model applied in analysis.

2.3.2 Instrumental Variable Analysis

While family structure analysis mostly gave way for within-family analysis, one method began to emerge in small circles of sibship size literature to address challenges presented by independent variable endogeneity and selection bias. Instrumental variable (IV) analysis relies on a selected instrument that exogenously affects the observed number of siblings without influencing the outcome variable (educational attainment and cognitive ability). In regards to sibship size effects, several instruments have been proposed with varying results.

One popular instrument proposed is the event of multiple simultaneous births (twins). The theoretical justification of the twin birth instrument is that the unplanned birth of an additional child accounts for family planning decisions that lead to sample selection (Diaz and Fiel, 2020).

With the exception of Black et al. (2010), use of the twin birth instruments have often confirmed doubts of sibship size causality (Black et al., 2005a; Dayioğlu et al., 2009; Diaz and Fiel, 2020). However, there are several valid grounds for criticism of this technique. Öberg (2017a) challenges the underlying assumption that twin births always capture *unwanted* additions to the family, undermining the instrument's exogenous influence on family size. For instance, new parents who have twins may have desired (and therefore planned for) two children from the offset, but just not in such quick succession (Öberg, 2017a, pp. 9). To ensure the validity of the twin births instrument, the empiricist would have to condition the instrument as a random event not linked to the desired number of children, requiring rather restrictive criteria difficult to achieve with smaller samples. Another criticism of the twin birth instrument is that there are possible biological differences between mothers who have twins and mothers who do not. For one, multiple previous pregnancies increases the probability of having twins (Ölberg 2017a, pp. 9). Thus, mother's who choose to have larger families are possibly overrepresented in twin birth samples. Secondly, Farbmacher et al. (2017) paradoxically argue that, since older mothers are more likely to have dizygotic twins, and high-income parents have a higher probability of delaying childbirth (Kreyenfeld, 2002), they are at higher risk in birthing twins. This violates the assumption once again that there are no systematic differences biologically or culturally between families with and without twin births. Öberg (2021) therefore urges the reevaluation of the use of the twin birth instrument in favor of alternatives.

One such alternative is the use of gender composition as an instrument. Use of this instrument lies on the assumption that most parents prefer a balanced gender composition and thus presence of multiple same-sex children increases the likelihood of parents to have another child (Angrist and Evans, 1996). While this was not originally intended to study family size effects on educational outcomes, Conley and Glauber (2006) indeed find that utility of this instrument leads to results consistent with previous sibship size literature. Conversely, Black et al. (2010) found that using the same instrument yielded no negative effect of sibship size effects. However, while still an improvement from twin birth instruments, gender composition instruments have similarly restrictive sampling criteria that limit the generalizability of the findings. Accordingly, restrictions result in the exclusion of households with (1) less than two parents, (2) less than two children, (3) opposite sex siblings, and (4) non-biological children—an extensive but no exhaustive list by the standards of Conley and Glauber (2006). Furthermore, unlike the twins instrument, the gender composition instrument does not result in an unplanned increase in family size. This has two relevant implications. First, if parents have the opportunity to prepare the birth of an additional child, this might mitigate proposed effects of resource allocation. Second, although presence of two same-gender children is a random event, non-random differences between parents that affect their ability to prepare for a third child might also influence educational outcomes of their child.

With respect to the drawbacks of previous IV-studies, the present thesis follows the methods of Jæger (2006) and Radyakin (2007) by employing multiple rather than single instruments. These instruments include: number of mother's siblings (MS), number of father's siblings (FS), mother's age at first birth (MA), and father's age at first birth (FA). The theoretical background of these instruments is rooted in demography research. It follows that the instrument qualifies as a natural experiment since partners are randomly assigned to each other with previously unknown "reproductive capabilities" (pp. 5). The number of siblings each parent has informs both fertility chances of parents based on familial cultural norms based around reproduction in which they were raised in addition to their genetic propensity to have children. While it could be argued that MS and FS violate the exclusion restriction principle of IV—i.e. that they also influence the dependent variable—this would only be the case if their respective aunts and uncles were actively present in the child's household. Similarly, the age of each parent at first birth may impact the educational outcomes of their children if we consider the theoretical assumptions of RDT; older parents may be more equipped than younger parents to provide their children with resources to aid them in their academic abilities. Both Jæger (2006) and Radyakin (2007) test these assumptions empirically and find no violation of the exclusion restriction principle. The present thesis also tests these instruments for robustness in chapter four.

Results from both Jæger (2006) and Radyakin (2007) have both indicated empirical support for the causal relationship between family size and educational outcomes. In fact, while additional siblings under conventional OLS regressions predict a one-tenth reduction in school years, Jæger (2006) finds that the use of this instrument yields a negative prediction of one-third of a school year per additional sibling. Similarly, Radyakin (2007), who considers the impact of sibship size on the probability of attending *gymnasium*, finds that IV analysis yields coefficients that are three times as large as that of OLS regressions (pp. 17). Derived from these results, the present thesis formulates the following hypothesis:

H1: Sibship size and educational attainment are negatively correlated even when independent variable endogeneity is controlled for.

Beyond simple replication, however, the aim of this thesis is to expand on their results, by considering the mechanisms of RDT and CT respectively. While it is possible to employ simple OLS regressions in line with Downey (1995), it must be determined whether the chosen instruments follow the exclusion restriction principle when individual mechanisms are treated as the dependent variable.

As discussed in section 2.2.1, parents allocate categorically distinct types of resources to their children, therefore theoretical arguments of whether MS, FS, MA, or FA violate exclusion restriction depend on the chosen mechanism. To repeat a previous argument, the number of siblings each parent has should only impact resource allocation if they are consistently involved (see Shavit and Pierce 1991) as they may help compensate for time and financial resources. If involvement of extended family is controlled for, however, violation of exclusion restriction based on the instruments MS and FS have no real theoretical standing. The age of each parent at first birth on the other hand demands a stricter evaluation. Older parents may have a higher income, more accumulated wealth, more cultural knowledge, and even more time. However, more important is whether this is consistently true. A series of robustness tests (i.e. Hansen J statistics) are conducted for different times of resources in chapter four.

With the confirmation that the instruments are not correlated with the error term of their respective dependent variables we derive the following hypothesis from RDT:

H2: Larger sibships are correlated with the dilution of both interpersonal and cultural resources in isolation of family size endogeneity.

Lastly, IV-analysis will be used to empirically test the validity of confluence theory. One concern in this test is the endogeneity of family structure variables such as age spacing and birth order. Yet, Jæger (2006) neither excludes these family structure characteristics nor claims that their IVs correct for their endogeneity. It is unclear whether this is an oversight or a feature in their analysis. For this reason, we carry out this analysis with caution. With this disclosure, I hypothesize the following:

H3: Larger sibships with covariance of family structure variables is negatively correlated with cognitive ability with the application of instrumental variables.

Chapter three: Research Design and Methods

3.1 Data

To replicate and expand on Radyakin's (2007) analysis, I follow suit by utilizing the youth questionnaire of the German Socioeconomic Panel Data (SOEP). Implemented into the larger panel in the year 2000 and repeated each year since, the youth dataset samples 17 year-olds already nested in households participating in the broader panel survey. Respondents are asked a broad range of questions about their home-life, educational achievements, ambitions, and their day-to-day hobbies. Although this is a pooled-cross sectional survey within the SOEP, the household structure of the panel design is advantageous in providing data on parental resources, cognitive ability, sibling constellations, and the chosen instruments of parental siblings and parental age at first birth. Moreover, the sample size has expanded to over 9,000 respondents. This alone allows the opportunity to test the reliability of Radyalin's (2007) findings. Beyond this, however, the motivation of this master's thesis is to expand on the scope of his analysis by testing specific mechanisms posited by RDT and to test the effects of sibling spacing and birth order in accordance to CT. The youth questionnaire data is ideal for this expansion.

3.2 Variables

3.2.1 Dependent Variables

Educational achievement measures are wide-ranging in the SOEP. In the youth survey, participants are asked to report grades in math, German (as a native language), and a foreign language. Furthermore, international classifications such as *comparative analysis of social mobility* (CASMIN) is also an available measure in the SOEP data. However, in line with Radyakin (2007) and Grätz (2018), I use attendance of the highest-tier secondary school (*gymansium)* as the primary outcome variable for educational attainment. Reasons for using this measure are twofold. First, family of origin effects play a stronger role on educational outcomes that occur at a younger age (Schneider, 2008). Thus, measures of later-life attainment are less relevant to the effects of sibship size, parental education, and parental socioeconomic status. Second, while there is variation between federal states, early tracking in Germany is often characterized by rigidity. Until recently, there was very little opportunity to move between tracks and obtain a tertiary degree. Secondary school sorting is often cited as a source of social reproduction in Germany and is therefore the most relevant outcome to this thesis. For a lengthier discussion of this decision see section 1.2.2 on the German context.

Cognitive ability is measured using the results of a cognitive test (*cogdj*) administered to some of the 17-year-old youth respondents from 2006 onward. The test measures cognitive mechanics in areas of verbal, numeric, figural ability (Richter et al., 2017). This was not included in Radyakin's (2007) analysis since the test was not integrated into the SOEP until 2006. I choose to include this variable as a secondary outcome variable as it has more theoretical relevance to confluence theory (CT). In accordance with Jæger's (2009) model, cognitive ability can be treated as a mediator of sibship size on educational attainment under both CT and RDT, whereas the direct effect of sibship size is exclusive to RDT. The inclusion of this outcome is used to test sibling constellation variables and CT. However, since the *cogdj* dataset is a truncated subsample of the youth survey, this is treated as a supplementary analysis not directly comparable to results of the primary analysis.

In addition to educational attainment and cognitive ability, the present thesis treats resources as dependent variables in order to empirically test resource dilution in the vein of Downey (1995). This provides additional context to the relationship between sibling configuration and educational attainment purported by the literature. In this regard, parental resources are categorized by *environment, opportunities,* and *treatment*.

Environment, as previously discussed, focuses on the cultural objects and educational activities of the home. By this understanding, the household environment is operationalized by measures of privacy (such as whether the youth participant has their own room) and educational activities (frequency of reading versus frequency of watching TV).

Opportunities are measured by participation in cultural activities that take place outside the home. These include participation in sports, music, dance/theater, and technical work such as computer programming.

Treatment is operationalized based on their child's subjective relationship with them as well as their engagement in their child's educational life. Some variables used to measure their child's *relationship* to their parents include their child's perceived importance of their mother/father in their life, the extent to which they talk with their respective parents about their personal experiences, and the degree to which their respective parents involve them in problem solving.

Parental involvement is operationalized by a series of variables that indicate their participation in their child's school life (i.e. attendance of parent meetings, helping with homework). For an exhaustive list of variables used for the quality of parent-child relationship and parental engagement see table 3.4 in the appendix.

3.2.2 Independent Variables

The primary independent variable of interest to the present master's thesis is sibling configuration. Sibling configuration is operationalized by family size, sibling spacing, and birth order.

Family size is measured by the number of children merged from the birth history of the respondent's mother. Simple OLS models will consider these numbers at their crude value, while IV models will adjust sibship size using the instruments discussed in the following section.

Sibling spacing (i.e. sibling density) is operationalized in a variety of ways by other scholars. Powell and Steelman (1990) construct a variable for *near siblings* (number of siblings within two years of age) and *far siblings* (number of siblings 3 years older or younger than the respondent). Additionally, the authors generate a measure that accounts for directionality—that is, siblings 1-2 years older etc.. Since we apply an instrument that affects sibship size these sibling density measures are unusable in my final models. Instead, I operationalize *age spacing* in line with Jæger (2006), by measuring the age difference in years between the respondent and next closest sibling (regardless of direct).

Birth order is found in the *biosib* dataset as a continuous variable where 1 is highest birth order and subsequent *n* represent lower birth order. To simplify the analysis, I again follow the operationalization of Jæger (2006), generating dummy variables for youngest, middle, and oldest children.

In addition to sibling configuration variables, both individual-level and household-level controls are added to the model.

Parental Socioeconomic status is measured by the highest score international socioeconomic status index (ISEI) between either parent. Since cases of parental separation are excluded in my analysis, it is unnecessary to consider both parents occupational prestige on the theoretical grounds of social reproduction. This decision also diverges from that of Radyakin (2007) who measures only the father's occupational prestige using a blue-collar and white-collar framework which assumes a male-breadwinner model and operates under an oversimplified and outdated classification. The current thesis opts for ISEI to account for these problems.

Parental education is measured by each respective parent's reported number of years in school or university. Alternatively, supplemental models measure parental education as a dummy variable whereby a value of 1 represents that a parent has an upper secondary school-leaving degree (*abitur*) and 0 indicating otherwise in line with the operationalization of Grätz (2015, 2018). Since use of this operationalization does not yield dramatically divergent results, they are not used in the main analysis.

Household income will also be considered as an additional control for family differences since material resources are relevant to resource dilution. Household income is measured by the year that the participant took the youth survey and is transformed using a natural log.

Immigration status is operationalized by whether the participant is a first generation immigrant (e.g. born outside of Germany) or second generation (child of at least one parent born outside Germany). This is an important control since cultural reproduction is contingent on knowledge and participation in the dominant culture from which immigrants are often excluded in the German context (Becker, 2010).

Gender of the survey participant is also controlled for. In the past, women were excluded from higher education. In recent studies, however, and reflected in SOEP data, women outperform men in academic performance and the accumulation of human capital. Furthermore, gender dynamics play a role in resource dilution (see section 2.1.2).

3.2.3 Instrumental Variables

As detailed in the previous chapter, the present master's thesis uses IV analysis to account for family size endogeneity. The main instruments include the number of siblings each parent has and the age of each parent at first birth. Information of the number of siblings for each parent are only asked in 1991, 2001, 2003, 2006, and 2011. Therefore, the original sample is truncated to respondent's whose parents do not have missing values in these survey years. If the values given between survey years differ, the highest value is chosen to estimate the true number of parental siblings. Additionally, age of parents at first birth is imputed by subtracting each parent's year of birth respectively by year of birth of the eldest child.

3.3 Methods of Analysis

The empirical goals of the present thesis are twofold: (1) to measure the extent to which use of instrumental variables changes sibship size effects, and (2) to understand the extent to which dilution mechanisms explain these effects.

3.3.1 Benchmark OLS Models

To accomplish these goals, OLS models are constructed to account for conventional measures of sibship size effects on educational attainment, cognitive ability, and the dilution of various resources.

$$PR(Gymnasium) = \beta_0 + \beta_1 Sibnum + \beta_2 X + \varepsilon \quad (1)$$

The first equation represents the benchmark model to measure biased OLS estimates of attending gymnasium against instrumental variable estimations. Since the probability of attending gymnasium is represented by a dummy variable, results are interpreted using linear probability. *Sibnum* represents the primary independent variable of interest, a generated measure of the respondent's number of sisters and brothers combined. A number of observable household and individual-level controls are represented by *X*. These include parental ISEI, parental level of education, and participant gender.

$$CogAbility = \beta_0 + \beta_1 Sibnum + \beta_2 SibDens + \beta_3 BirthOrder + \beta_4 X + \varepsilon \quad (2)$$

Similar to equation (1), equation (2) predicts cognitive ability using linear regression to yield biased estimates to compare against the effects of IV analysis. *Sibnum, SibDens,* and *BirthOrder* represents measures of sibling configuration measured by variables of sibship size, age spacing, and birth order. In line with Jæger (2006), these measures are run together rather than discreetly in accordance with confluence theory. Controls represented by X will also be used in models to account for between family differences.

$$Resources = \beta_0 + \beta_1 Sibnum + \beta_3 X + \varepsilon \quad (3)$$

Equation (3) represents OLS estimates of the dilution of resources with respect to unstandardized family size (*Sibnum*) and various controls represented by X. This is the third benchmark model specifically aimed at reaching the second empirical goal of the present thesis. Separate models will measure the dilution of different types of resources using linear regression since resources are constructed using their own distinct indices.

3.3.2 Instrumental Variable Analysis

As discussed in the previous section, in each of these models the main explanatory variable (*Sibnum*) is correlated with the error term represented by ε. Therefore, multiple instrumental variable analysis is introduced in our second set of models to standardize the endogenous character of family size.

$$ISibnum = \alpha_0 + \alpha_1 MS + \alpha_3 FS + \alpha_4 MA + \alpha_5 FA + \alpha_6 X + \delta \qquad (5)$$

The first step of controlling for heterogeneity between families is shown in equation one. Instruments of number of mother's siblings (MS), number of father's siblings (FS), mother's age at first birth (MA), and father's age at first birth (FA) are used to impute an instrumented value for number of siblings (*ISibnum*).

$$PR(Gymnasium) = \beta_0 + \beta_1 ISibnum + \beta_3 X + \varepsilon \qquad (6)$$

$$CogAbility = \beta_0 + \beta_1 ISibnum + \beta_2 SibDens + \beta_3 BirthOrder + \beta_4 X + \varepsilon \qquad (7)$$

$$Resources = \beta_0 + \beta_1 ISibnum + \beta_3 X + \varepsilon \qquad (8)$$

The second step of applying instrumental variable analysis is shown in equations 6-8, wherein the instrumented number of siblings is applied to the original OLS models. Instrumented sibship size (*ISibnum*) should no longer correlated with the error term ε and therefore yield unbiased estimates. Application of two-step instrumental variable analysis is achieved with the command *ivregress 2SLS* in version 18 of Stata SE.

3.3.3 IV Diagnostic Tests

In light of the expansion of Radyakin's (2007) analysis conducted in equations 7 and 8, it is especially important to test the validity of our instruments. Specifically two criteria must be met: (1) the instruments must be correlated with the endogenous variable—an assumption already tested in source literature and (2) the instruments cannot be correlated with the outcome variable (known as exclusion restriction principle). Testing of exclusion restriction principle is particularly important in my extended analysis, because I introduce two previously untested outcome variables in cognitive ability and parental resources. Two diagnostic tests are conducted to ensure these criteria are met.

An F-Statistic test is conducted by examining the results of the first stage of *2SLS* represented in equation 5. This test is used to determine the empirical relevance of the chosen instruments. If the F-Statistic is greater than 10, this is an indication of a strong instrument that is correlated with the endogenous variable (Stagier and Stock, 1994). Both Jæger (2006) and Radyakin (2007) conduct this test and find a strong F-Statistic. I repeat this test and find similar results.

A Hansen-J statistic and its corresponding chi-squared p-value are yielded from the results of each model. The null hypothesis of a Hansen-J statistic is that the overidentification restrictions are valid--that is, the instruments are not correlated with the error term. Yielding a small p-value leads to a rejection of the null hypothesis, meaning that at least one instrument is correlated with the error term. It is not possible to test directly whether an instrument is jointly correlated with the outcome variable, however, a rejection of the null hypothesis provides evidence for this possibility and invalidates any estimates yielded from the model. As discussed in chapter four, a number of the models tested and replicated yield small p-values (see table 4.17 in the appendix). The implications of this finding are discussed at length in the conclusion.

Chapter Four: Empirical Results

4.1 Probability of Gymnasium Attendance

One of the primary goals of the present thesis was to replicate the findings of Radyakin (2007) using an updated pooled sample. In the following section, the selection procedure is described including key decisions made that differ from that of Radyakin. Thereafter, the composition of the sample will be compared by track attendance. Multivariate analysis is then applied to this updated sample first using linear probability models of ordinary least squares (OLS) and then instrumental variable (IV) analysis. Finally diagnostic tests are used to test the validity of these instruments.

4.1.1 Descriptive Statistics and Bivariate Analysis

To generate an appropriate dataset, I start with the 9,739 respondents that filled out the youth questionnaire between 2000 and 2021. This data is then merged with variables related to parental education and parental occupational class from the *bioparen* data sets. Supplemental information about their parents used to generate the instrumental variables are then merged from the individual questionnaire filled out by their parents in the years this question is available. These datasets are merged by matching their parent's identification numbers (*pid*) with their respective parental indicators found in bioparen (*fnr mnr*). Additional information on the youth respondent such as migration status, later educational attainment, birth year, and gender are merged from other datasets. Finally, household income from the year the survey was filled out was merged from the household generated dataset (*hgen*).

Radyakin (2007) excludes respondents who did not live with both biological parents for the first 15 years of their childhood i.e. until the survey was administered. The explanation for this is twofold: (1) both parents must be present so that information on the instruments is available and (2) parental separation may have additional effects on educational attainment separate from that of family size. Since school tracking occurs at age 10 or 11, previous scholars who have considered the effects of parental separation have selected respondents who were not living with both parents before this age (Grätz 2015). However, since my secondary analysis and tertiary consider outcomes at the age the respondent filled out the survey, I also exclude respondents who do not live with both biological parents. Additionally, I exclude observations where information one or both parents is missing.

The primary outcome variable, school track attendance, is directly asked in the youth dataset. However, specifications on which track the respondent attends is only available from the year

2006. In order to append missing data before 2006, I merge proxy data on school-track attendance from the *kidlong* survey filled out by their parents in the years before the respondent turned 16. I take the maximum value between the two datasets, then I generate a dummy variable on whether the respondent attended gymnasium.

The final dataset results in 3,993 observations with youth participants born between the years 1984 and 2000. Eliminated from the sample are respondents who filled out the survey after the year 2018 since no matching parental data is found from the *bioparen* dataset.

Table 4.1 presents the mean values of each variable relevant to our analysis sorted by school track attendance based on our imputed measure. Overall, approximately 46% of our sample attended the highest track (*Gymnasium*) or acquired a high-level school leaving degree (*Abitur*), while the remaining 53% attended lower school tracks or did not go on to acquire a qualification to attend university.

Most consequential to our research question is whether smaller families are more likely to attend the highest school track. From table 4.1, respondents from gymnasium have approximately 0.22 fewer siblings than respondents who attended other tracks. This figure is only marginally smaller than the 0.3 family size difference reported by Radyakin (2007), indicating that OLS regression analysis will yield somewhat similar results.

Women are slightly overrepresented in the gymnasium track compared to men despite being underrepresented in the sample overall. Since this is a pooled sample from different survey years, age is included in our analysis to control for cohort effects. As seen in the table, pupils who attend gymnasium are slightly older. This could indicate that admission to gymnasium became gradually less common between survey years.

Respondents with direct and indirect migration background account for 23% of respondents. Nevertheless both first generation and second generations immigrants are underrepresented in the gymnasium track compared to other tracks.

Parental years of education are included to account for social reproduction theory discussed in the first chapter. Father's have on average more education than mother's in this sample, but differences between years of education between the mother and the father is largest in the gymnasium track (0.49 years). This indicates that the father's education may have a larger influence on educational attainment than that of the mother. In either regard, there is a clear indication that respondents who attended gymnasium have more educated parents on average than do children of other tracks. Father's of respondents who attended gymnasium have on average 2.4 more years of education than Father's of respondents who attended other tracks. These figures are comparable albeit slightly smaller than those found by Radyakin (2007). Bolstering these figures, are reports that parent's holding a higher school-leaving *abitur* degree have children overrepresented in gymnasium tracks.

As discussed in the previous chapter, rather than reporting father's occupational and socioeconomic status, I take the highest occupational status reported between parents. Additionally, I use occupational status as Erikson-Goldthorpe-Portocarero (EGP) class scheme to illustrate class differences in upper-track attendance. Pupils who attended upper-track schools are more likely to have parents who work in Class-I (higher managerial and professional works) or Class-II (lower managerial and professional works) as indicated by the mean of a dummy variable. Table 4.1 also reports the mean for each individual class by track attendance. Overall, representation decreases with each lower class. However, since I am dealing with a relatively smaller sample size with very few cases for lower classes, I choose to exclude EGP in my regression analysis in favor of a ISEI linear index. Descriptive statistics indicate that pupils who attend gymnasium have at least one parent with a SEI score about 14 points higher than that of pupils from other tracks.

Lastly, income seems to be an indicator of track attendance. Participants who were sorted into the gymnasium track come from wealthier families—with an approximate 37.71% increase in household income compared to pupils of other tracks.

4.1.2 Biased OLS Estimates

Table 4.2 reports results from three OLS models *Gymnasium* attendance as the dependent variable. Model 1 reports a family size coefficient that is both negative and statistically significant with the inclusion of controls for the respondent's gender, age cohort, and immigrant status. Accordingly, having an additional sibling is correlated with decreasing probability of attending *Gymnasium* by 3.5 percentage points. Being female increases the probability of upper-track attendance by about 5.4 percentage points. The respondents age holds significance and decreases the probability of upper-track attendance by 1.44 percentage points.

Effects of migration status are heterogenous. Being a first generation immigrant is negative but not statistically significant in the first model, but increases in effect size and significance in later models. Inversely, second generation immigration status is very highly significant in the first model but is washed out with the introduction of other controls.

The second model introduces parental education variables and household income to account for effects of social reproduction. The first glaring observation is that family size remains significant and only slightly decreases in effect size. In light of this, however, parental education and income

Figure 4.1 OLS results for probability of attending Gymnasium.

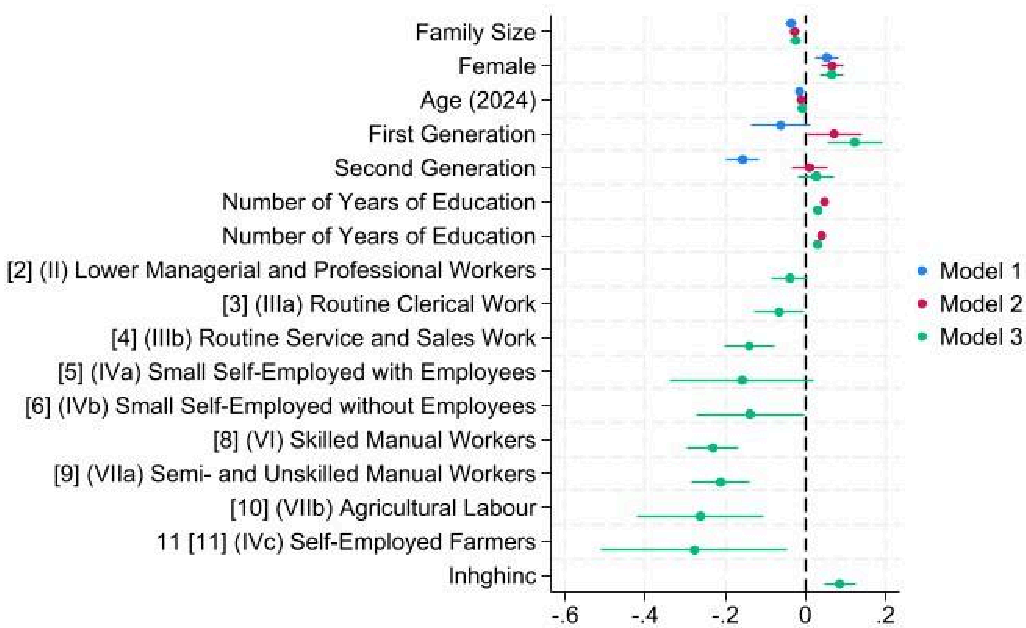


Table 4.2 OLS regression estimates for Gymnasium attendance.

	(1) OLS1	(2) OLS2	(3) OLS3
Family Size	-0.0359***	-0.0273***	-0.0274***
	(0.000)	(0.000)	(0.000)
Female	0.0536***	0.0664***	0.0678***
	(0.001)	(0.000)	(0.000)
Age (2024)	-0.0144***	-0.0105***	-0.00863***
	(0.000)	(0.000)	(0.000)
First Generation	-0.0615	0.0723*	0.107**
	(0.103)	(0.039)	(0.002)
Second Generation	-0.157***	0.0103	0.0121
	(0.000)	(0.649)	(0.586)
Number of Years of Education (Father)		0.0481***	0.0298***
		(0.000)	(0.000)
Number of Years of Education (Mother)		0.0407***	0.0320***
		(0.000)	(0.000)
Parental ISEI			0.00426***
			(0.000)
Income (Logged)			0.0876***
			(0.000)
Constant	1.028***	-0.276***	-0.927***
	(0.000)	(0.000)	(0.000)
Observations	3814	3814	3814

p-values in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

are both highly significant and have a much larger effect on the probability of upper-track attendance. Also notable in this model is that father's education has only a marginally larger effect on the upper-track attendance than mother's education. This is a much smaller difference than observed by Radyakin (2007), and the pattern inverts itself in the full model.

The third model adds parental ISEI and logged income as control variables. All variables are highly significant in the final model with the exception of second generation migration status. As expected, both income and parental ISEI are influential on the upper-track attendance. A one percent increase in income is correlated with a .0876% increase in probability, while mobility of only one point higher on the ISEI index is correlated with a .426% increase in probability of attending gymnasium. While these estimates seem trivial, it is important to remember that they represent a very granular scale. To illustrate the effects of class on upper-track attendance, I present the final model using EGP as a measure of occupational prestige in figure 4.1 in which estimates gradually creep further left on the x-axis, representing ever-decreasing chances of upper-track attendance with lower social class. However, important to note is that confidence intervals widen, with certain EGP groups due to smaller sample sizes in these classes. For this reason, EGP is left out of the final model and only presented here for illustrative purposes.

In summary, standard OLS estimates are very consistent with findings of Radyakin (2007) and other previous studies. Family size remains robust and statistically significant with each additive model even with different choices in operationalization. A few relevant deviations from his findings however are in the strength and significance of migration background controls, with first generation migration status decreasing the chances of upper-track attendance by 10.7%, outweighing the impact of all other independent variables. Radyakin's (2007) final model also reports a higher r-squared value of .33, with the independent variables in my final model only accounting for 24% of the variance. However, as previously established, standard OLS estimates likely introduce selection bias, yielding underestimated standard errors. For this reason, these estimates should only be taken as a benchmark to compare with IV regression analysis.

4.1.3 IV Regression Analysis

To adjust for sibship size endogeneity, we introduce the four instruments discussed in the previous section and apply two-step instrumental variable analysis. Table 4.3 presents each stage of IV-LPM regression for each of the three additive models discussed above using the binary measure of school track attendance. In the first stage of model 1, all instruments carry statistical significance except father's age at first birth, which remains insignificant in successive first-stage regression models. Predictably, mother's age at first birth is the only negative coefficient contradicting time-squeeze effects theorized by some demographers that educated women who are racing against their biological clock have more children in quicker succession.

Figure 4.2 IV results for probability of attending Gymnasium.

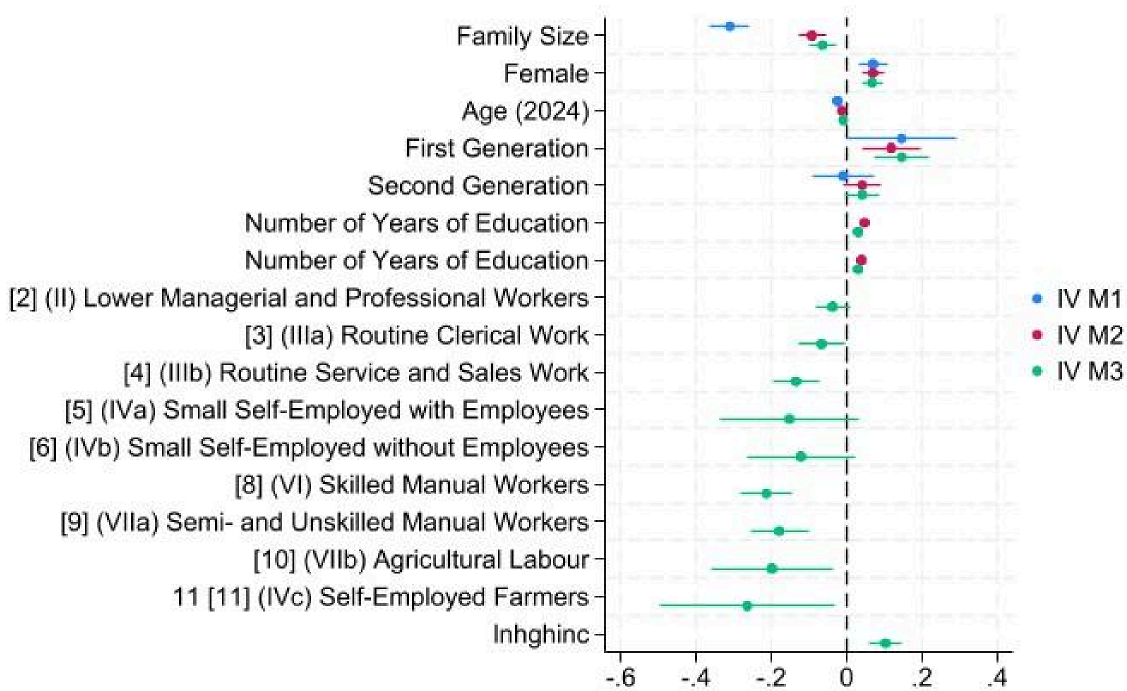


Table 4.3 First and second stage IV estimates for Gymnasium attendance.

	(1)	(2)	(3)	(4)	(3)	(3)
	IV1: First-stage regression	IV1: Second-stage regression	IV2: First-stage regression	IV2: Second-stage regression	IV3: First-stage regression	IV3: Second-stage regression
Family Size		-0.310^{***}		-0.0925^{***}		-0.0680^{***}
		(0.000)		(0.000)		(0.000)
No. of Siblings (Mother)	0.0750^{***}		0.0853^{***}		0.0836^{***}	
	(0.000)		(0.000)		(0.000)	
No. of Siblings (Father)	0.0701^{***}		0.0810^{***}		0.0799^{***}	
	(0.000)		(0.000)		(0.000)	
Mother's Age (first birth)	-0.0783^{***}		-0.0904^{***}		-0.0916^{***}	
	(0.000)		(0.000)		(0.000)	
Father's Age (first birth)	0.00469		0.00350		0.00262	
	(0.523)		(0.628)		(0.701)	
Female	0.0456	0.0699^{***}	0.0531	0.0701^{***}	0.0441	0.0696^{***}
	(0.204)	(0.000)	(0.137)	(0.000)	(0.212)	(0.000)
Age (2024)	-0.0550^{***}	-0.0242^{***}	-0.0558^{***}	-0.0129^{***}	-0.0448^{***}	-0.00969^{***}
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
First Generation	0.423^{+}	0.146^{*}	0.474^{*}	0.119^{**}	0.478^{*}	0.135^{***}
	(0.055)	(0.048)	(0.032)	(0.002)	(0.023)	(0.000)
Second Generation	0.203^{+}	-0.00994	0.275^{*}	0.0415	0.250^{*}	0.0301
	(0.072)	(0.811)	(0.017)	(0.103)	(0.025)	(0.210)
Father's Edu (yrs)			0.0402^{***}	0.0479^{***}	0.0229^{*}	0.0298^{***}
			(0.000)	(0.000)	(0.041)	(0.000)
Mother's Edu (yrs)			0.0250^{*}	0.0391^{***}	0.0181	0.0311^{***}
			(0.050)	(0.000)	(0.167)	(0.000)
Parental ISEI					-0.00500^{+}	0.00380^{***}
					(0.062)	(0.000)
Income (logged)					0.498^{***}	0.107^{***}
					(0.000)	(0.000)
Constant	5.967^{***}	2.059^{***}	5.448^{***}	-0.00558	1.682^{*}	-0.905^{***}
	(0.000)	(0.000)	(0.000)	(0.958)	(0.018)	(0.000)
Observations	3814	3814	3814	3814	3814	3814

p-values in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Both maternal and paternal sibship size are positively correlated with the family size of the respondent, however, consistent with previous findings, the number of siblings of the respondents mother is more influential that of their father. This pattern and the predictive power of the instruments persists with each additive model.

Beyond the instrumental variables, there seems to be a significant cohort effect on family size, with an increase in age by one year in 2024 correlating with the reduction of .05 siblings, indicating possible demographic shifts in family size over time that requires further inspection. Lastly, both first and second generation immigration status has a large influence on family size albeit with very small statistical significance that increases in later models.

The second model once again introduces parental education as control variables. Both parent's level of education seem to have a positive effect on family size with the inclusion of the instruments, but this effectively disappears once controlled for income and parental ISEI in the last first-stage model. Parental education also remains a predictive measure of gymnasium attendance in the second-stage of this model. Consistent with the OLS estimates, father's education matters more than mother's education until the pattern is reversed with the introduction of additional controls.

The third model accounts for logged income and parental ISEI and generally yields results consistent with OLS estimates. Parental socioeconomic status negatively predicts family size when instrumental variables are for, while logged income is strongly and significantly positively correlated with family size in first-stage regression. This implies that when cultural conventions of family planning are accounted for by the chosen instruments, income significantly increases the chances of having larger families. Unsurprisingly both household income and socioeconomic status remain strong predictors of upper-track school attendance with instrumented family size.

The most consequential finding of IV linear probability models is the instrumented effect of family size on school-track attendance. Similar to the findings with Radyakin (2007) instrumental variables increase the effect of family size by a significant magnitude, although to a lesser degree. Instrumental estimates of family size are in fact about 2.5 times larger than that of OLS, compared to the reported factor of 3 reported by Radyakin (2007). Unlike findings from other instruments, these results support evidence that causal effects of family size are rather underestimated in conventional OLS analysis. However, the weight of these findings relies heavily on the validity of their instruments.

Jæger (2006) argues that the benefit of multiple instruments is that their validity is empirically testable compared to use of single instruments. Both Jæger (2006) and Radyakin (2007) rely on F-Statistic and Anderson canonical correlations to reject underidentification of the instruments. IV estimates in model three of table 4.3 yield an F-Statistic of 79, well above the rule of thumb of 10—indicating that like previous findings, the instruments are relevant to the model and identify the endogenous variable. Previous studies also exploit the Sargan-Hansen test of overidentifying restrictions to test if the instruments are correlated with the error term. Since my analysis clusters results by household to account for intercorrelation between siblings in my data, I rely on the Hansen J statistic as an alternative diagnostic test. Unfortunately, unlike previous scholars, all three IV models yield a p-value smaller than 0.05, resulting in a rejection of the null hypothesis that the instruments are uncorrelated with the error term. This is an indication that at least one of the instruments is correlated with the upper-track attendance, violating the assumption of restriction exclusion.

Kreyenfeld and Konietzka (2008) identified key differences in fertility patterns between mother's with different levels of education, finding that women with more education tend to delay motherhood well into their thirties. From this theoretical perspective, it seems very probable there is an unmeasured difference between older and younger mothers that impacts their child's educational chances. To test this hypothesis, I reran the models three times over, excluding mother's age at first birth, father's age at first birth, and both respectively—finding that when both instruments are excluded from the model, the Hansen-J statistic yields a p-value well above 0.05. While this supports my hypothesis, further evaluation is needed to assess the validity of these instruments. A full report of these diagnostic tests is found in table 4.17 of the appendix.

In this section, I report the findings of both OLS and IV regression models to test the first hypothesis of the present master thesis:

H1: Sibship size and educational attainment are negatively correlated even when independent variable endogeneity is controlled for.

While estimates from both conventional OLS and IV estimates of family size are consistently negative and highly significant in every model, diagnostic tests of the chosen instruments indicate that estimates are compromised by overidentification. In this regard, in dissent of previous literature, my findings do not yield enough evidence to confirm this hypothesis.

4.2 Testing Resource Dilution

While evidence for instrumented family size effects on upper-track attendance is inconclusive, there is still theoretical grounds for empirically testing the actual dilution of resources proposed by RDT. This not only provides an updated examination of work done by Downey (1995), it permits further testing of the validity of the instruments utilizing other dependent variables.

4.2.1 Descriptive Statistics and Bivariate Analysis

To test RDT more directly, I start with the youth sample and apply a similar procedure of data management and selection detailed in the previous section. This includes matching identification variables from *bioparen* to individual datasets to scavenge information related to parental controls and the four instrumental variables. Once again, I exclude cases of parental separation or where information on at least one parent is missing. Where my sampling procedure differs from the previous analysis is the inclusion of cases in which school-track attendance is missing. My justification for this is twofold. First, gymnasium attendance is no longer a relevant dependent variable. Instead the goal of these models is to investigate resource dilution more generally, rather than as a function of gymnasium attendance specifically. Second, exclusion of this outcome variable ensures a larger sample size in the case of missing values in the mechanisms investigated.

As discussed in the previous chapter, the youth dataset provides rich information on the respondents interests, relationship with their parents, and their perception of their parent's involvement in their academic performance. I exploit these variables as proxies for the three types of resources parents provide that may disintegrate with increasing family size. I first recode the variables so that higher values indicate greater availability or exploitation of these resources. For example, questions asking about the frequency of cultural activities such as reading and making music are reverse coded where a high value indicates higher frequency. Lastly, for the sake of comparability and simplicity of analysis, I exclude all missings of the dependent variables.

Tables 4.4 in the appendix reports the mean response to questions about resources related to home environment and cultural activities. For illustrative purposes, family size is re-categorized by single child families, two children families, three to four children families, and families with five or more children respectively. As expected, the mean value of whether a respondent has their own room gradually increases with family size, indicating that pupils from larger families may be less likely to have their own room in the home. Activities of reading and TV are also indicators of home environment, but do not follow the expected pattern of resource dilution. Frequency of watching television is considered an activity that is less conducive to learning and education. Therefore, following RDT, television viewing should increase in larger families. Mean values in table 4.4 indicate the inverse pattern. Unlike television viewing, frequency of reading indicates a richer home environment with a larger presence of educational material. Frequency of reading seems to follow a curvilinear pattern by sibship size, initially decreasing in two children families, increasing again in three to four children families, and then once again dropping off in larger families. Opportunities are measured by cultural activities that take place outside the house. Sports, music, and dance do not follow a clear linear pattern and differences between means are trivial. Frequency of working on technical work, such as computer programming, on the other hand gradually decreases with increasing family size.

Table 4.5 in the appendix presents bivariate results of parental treatment by family size, particularly the respondent's perceived relationship with their parents. The original variables measured the relationship the youth respondent had with each parent, since resource dilution theory does not focus on gendered division of parenting, I take the mean response between questions about each parent. Mean values of whether the parents are important in their life does not follow any clear linear pattern and is likely not correlated with family size. Indicators of parental attention such as whether they ask before making a decision, show appreciation, have the impression of trusting the respondent, give reason for making a decision, and show their child love decreases somewhat linearly with increasing family size. There seems to be an only child disadvantage for some resources, namely whether a parent talks to the respondent about their life and whether their parents involve their child in problem-solving. However, the differences between means of respondents from different family sizes varies in magnitude.

Lastly, table 4.6 reports the mean response of resources related to parental involvement in their child's education. Parental interest in the respondent's educational performance and whether they help with studying seems to wane somewhat with increasing family size. Attendance of parent's evening at the respondent's school again follows a slight curvilinear pattern with the mean response rising between only children and two-child families and then dropping off in larger families. The remaining indicators of parental involvement do not seem to follow clear resource dilution patterns although the mean response is consistently lowest in families with five or more children.

While bivariate analysis provides some evidence for resource dilution, consistent with Downey (1995) a granular examination of individual resources of different types show that not all resources follow a clear linear pattern of increasing scarcity. Furthermore, mean differences between respondents from different family sizes do not yield dramatic results, even though pupils from the largest families always report fewer resources than those from the smallest. It is also important to note that allocation of resources is dependent on the families propensity to provide them. It is therefore necessary to apply multivariate analysis to investigate whether patterns exist controlling for family heterogeneity.

4.2.2 Biased OLS Estimates

Estimates using ordinary least squares (OLS) are reported in tables 4.7 through 4.9 for each resource type respectively. Each column presents a regression model treating each resource as a

Table 4.7 OLS estimates of dilution of environment and cultural opportunities.

	(1) Reading	(2) TV	(3) Has own room	(4) Sport	(5) Music	(6) Dance	(7) Tech Work
Family Size	**0.0441***	**-0.112*****	**-0.0373*****	0.00823	**0.0705****	-0.0225	**-0.0514*****
	(0.026)	(0.000)	(0.000)	(0.639)	(0.004)	(0.219)	(0.001)
Female	0.421***	-0.0860***	0.00135	-0.228***	0.156**	0.573***	-0.822***
	(0.000)	(0.001)	(0.893)	(0.000)	(0.004)	(0.000)	(0.000)
Age (2024)	0.0177***	0.00522^{+}	-0.00358**	-0.0225***	0.0253***	0.0129**	0.0223***
	(0.000)	(0.072)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
First Generation	0.408***	-0.194^{+}	-0.139***	0.137	-0.0260	-0.123	-0.196*
	(0.000)	(0.051)	(0.000)	(0.209)	(0.836)	(0.177)	(0.040)
Second Generation	0.0674	-0.114**	-0.170***	-0.0871	0.00156	0.0439	-0.0194
	(0.336)	(0.007)	(0.000)	(0.185)	(0.985)	(0.482)	(0.730)
Father's Edu (yrs)	0.0705***	-0.00721	-0.00294	0.0198^{+}	0.0721***	0.00829	-0.0113
	(0.000)	(0.345)	(0.294)	(0.068)	(0.000)	(0.462)	(0.253)
Mother's Edu (yrs)	0.0344**	-0.00960	0.00508*	0.0354***	0.0252^{+}	0.0315**	0.00227
	(0.005)	(0.174)	(0.041)	(0.001)	(0.094)	(0.005)	(0.818)
Parental ISEI	0.00442*	-0.00163	0.00152**	0.00300	0.00225	0.00213	0.00272
	(0.034)	(0.195)	(0.002)	(0.118)	(0.360)	(0.255)	(0.114)
Income (logged)	-0.0724	-0.0282	0.0477***	0.130*	0.188*	0.161**	0.0930^{+}
	(0.271)	(0.465)	(0.000)	(0.025)	(0.012)	(0.004)	(0.074)
Constant	1.386*	5.367***	0.667***	2.538***	0.0379	-0.551	0.996*
	(0.011)	(0.000)	(0.000)	(0.000)	(0.951)	(0.232)	(0.022)
Observations	3421	3421	3421	3421	3421	3421	3421

p-values in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

dependent variable with a full set of controls for gender, age, immigration status, parental education, parental ISEI, and logged income. Variables with binary outcomes such as whether the respondent has their own room are interpreted using linear probability, while other regressions are interpreted using simple linear regression. To illustrate the significance and effect of family size on each resource, figures 4.3 through 4.6 present coefficient tables for each mechanism net of control variables.

Indicators of environmental resources reading, TV, and privacy all yield statistically significant coefficients but are partially inconsistent with patterns hypothesized by RDT. Unsurprisingly the presence of an additional child decreases the chances the respondent has their own room by 3.74 percentage points. This could indicate that, as predicted by Blake (1981), children in larger families are less likely to have privacy to learn and reflect effectively. Immigration status, parental socioeconomic status, logged income also impact the probability of a respondent having their own room as expected. Mother's years of education seems to have a small and statistically significant positive effect on this outcome variable, while Father's education does not seem to matter. More unexpectedly are the effects of family size on reading and TV. As previously discussed, resource dilution theory predicts that family size should decrease activities valuable to cultural reproduction. Yet an additional sibling is correlated with an increase of frequency of reading and a decrease in television viewing. Reading also seems to be positively correlated with being female, parent's years of education, socioeconomic status, and immigration status. Television viewing, on the other hand, is negatively correlated by being female and immigration status with other covariates yielding no other significant results.

Only two indicators of opportunities yield significant results: music and technical work. An increase in family size has a positive correlation with frequency of making music. Being female, age, parental education and logged income also yield statistical significance and are positively correlated with frequency of creating music. Inconsistent with Blake's (1989) findings, the effects of family size on frequency of dance/performing arts and sports do not yield statistical significance. Family size is however negatively correlated with the frequency of technical work

Table 4.8 OLS estimates for dilution of parental treatment.

	(1) Parent Important	(2) Parent Talks	(3) Parent Consults	(4) Parent Appreciates	(5) Parent solves	(6) Parent trusts	(7) Parent asks opinion	(8) Parent gives reason	(9) Parent shows love
Family Size	-0.0109	**-0.0308***	**-0.0480*****	**-0.0297***	-0.00676	**-0.0216**$^{+}$	**-0.0732*****	-0.0227	**-0.0350***
	(0.121)	(0.013)	(0.001)	(0.015)	(0.617)	(0.085)	(0.000)	(0.166)	(0.015)
Female	0.0438**	0.0841**	0.116***	0.0577*	-0.0417	-0.0434	0.0708*	-0.0323	0.0345
	(0.008)	(0.003)	(0.001)	(0.024)	(0.171)	(0.117)	(0.032)	(0.327)	(0.209)
Age (2024)	-0.00795***	-0.00491^{+}	0.00271	-0.00964***	-0.0129***	-0.00462	-0.00729*	-0.0131***	-0.0157***
	(0.000)	(0.096)	(0.437)	(0.000)	(0.000)	(0.106)	(0.044)	(0.000)	(0.000)
First Generation	0.0756^{+}	0.0238	-0.249**	-0.0714	-0.0221	0.0149	-0.169*	-0.0957	0.153*
	(0.054)	(0.724)	(0.007)	(0.305)	(0.774)	(0.838)	(0.049)	(0.260)	(0.029)
Second Generation	0.0252	-0.0412	-0.0164	-0.0239	0.00317	-0.0667	-0.130*	-0.0569	0.0782^{+}
	(0.232)	(0.315)	(0.740)	(0.537)	(0.946)	(0.109)	(0.015)	(0.246)	(0.060)
Father's Edu (yrs.)	-0.00131	0.0103	0.0152^{+}	0.00838	0.00802	-0.00687	0.00118	0.0115	0.00349
	(0.760)	(0.162)	(0.093)	(0.229)	(0.327)	(0.342)	(0.896)	(0.180)	(0.643)
Mother's Edu (yrs.)	-0.0116**	0.0121^{+}	0.0210*	-0.00437	0.0128	0.00193	0.0128	0.0249**	-0.0139^{+}
	(0.007)	(0.087)	(0.015)	(0.516)	(0.111)	(0.783)	(0.141)	(0.003)	(0.055)
Parental ISEI	-0.000909	0.00108	0.00428**	0.00292*	0.00260^{+}	0.00140	0.00243	0.00491**	0.00114
	(0.212)	(0.384)	(0.005)	(0.016)	(0.068)	(0.282)	(0.114)	(0.002)	(0.371)
Income (logged)	0.0649**	0.121**	0.0171	0.0758^{+}	0.0177	0.121**	0.0762	0.0618	0.130**
	(0.007)	(0.002)	(0.720)	(0.068)	(0.690)	(0.002)	(0.134)	(0.205)	(0.001)
Constant	4.631***	2.494***	2.868***	3.540***	3.528***	3.406***	2.959***	2.871***	3.773***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3421	3421	3421	3421	3421	3421	3421	3421	3421

p-values in parentheses $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

such as computer programming. This is an activity previously unexplored in resource dilution studies but may have increasing relevance in social reproduction with the introduction of technical computer knowledge in the job market and field of education. Being female and first generation immigrant have a significant negative impact on frequency of technical work, while logged income unsurprisingly yields a positive correlation.

Table 4.8 reports regression results from indicators of parental treatment as dependent variables. As illustrated by Figure 4.4, five of the nine indicators yield statistically significant results net of control. Accordingly, family size predictably decreases the likelihood that the respondent's parents talk with them, consult them on decisions, express appreciation for the things they do, ask opinion on family matters, and express love. Gender has a consistent significant effect on these indicators, as being female is positively correlated with increasing frequency of the aforementioned indicators. Immigration also yields significant results for some of the aforementioned resources, namely direct immigration status is negatively correlated with the frequency of parental consultation and asking opinion on family matters, while having both direct and indirect immigration background is associated with a perceived increased expression of love from parents. Father's education seems to play no significant role in parental treatment. Mother's education, on the other hand, is slightly more important—negatively impacting the perceived importance of the parent and expression of love (albeit at a low significance level), while positively impacting frequency of consultation and giving reason for decisions. Parental socioeconomic status yields significant positive coefficients on frequency of consultation, appreciation, collaborative problem-solving (at .10 significance level), and giving reason for decisions made. Similarly logged income shows a positive correlation for perceived importance of parents as well as frequency of talks with parents and perceived expression of trust and love.

Lastly, table 4.9 presents OLS estimates of family size on the dilution of parental involvement in their child's education. Of the six dependent variables, family size only has significant negative impact on parental interest in academic performance, whether the parents help with studying, and whether the parent's take part in parent's evening at their child's school. Gender effects do not follow a consistent pattern. On the one hand, being female is positively correlated with parental interest and increase their probability of meeting their teacher outside of appointed conferences, but conversely negatively correlated with attending parent's evening. Immigration status is more consistent. Being a first generation immigrant—that is, being born outside Germany and also

Table 4.9 OLS estimates of dilution of parental involvement.

	Parental Interest	Parental Help	Parent's Evening	Parent teacher Conference	Meets Teacher	Parent Representative
Family Size	**-0.0493*****	**-0.0192****	**-0.0280*****	-0.00824	0.000843	0.00525
	(0.000)	(0.002)	(0.000)	(0.292)	(0.889)	(0.379)
Female	0.0811***	0.00641	-0.0302*	-0.0280^{+}	0.0495***	0.0166
	(0.001)	(0.633)	(0.032)	(0.097)	(0.001)	(0.216)
Age (2024)	0.00139	-0.00283*	-0.00198	-0.00164	0.00230	-0.00105
	(0.607)	(0.043)	(0.192)	(0.362)	(0.122)	(0.471)
First Generation	-0.155*	-0.0596	-0.149***	0.00818	-0.0889**	-0.136***
	(0.032)	(0.106)	(0.001)	(0.845)	(0.004)	(0.000)
Second Generation	-0.00498	-0.0873***	-0.0874***	-0.0225	-0.0365^{+}	-0.0854***
	(0.894)	(0.000)	(0.000)	(0.378)	(0.081)	(0.000)
Father's Edu (yrs.)	-0.0187**	0.00203	0.000780	-0.00664	-0.000504	0.00731^{+}
	(0.006)	(0.573)	(0.843)	(0.159)	(0.895)	(0.081)
Mother's Edu (yrs.)	-0.0106	0.00337	0.0144***	-0.00386	-0.00265	0.0126**
	(0.109)	(0.333)	(0.000)	(0.405)	(0.488)	(0.002)
Parental ISEI	0.00165	0.000480	0.000773	0.000551	0.000732	0.00144*
	(0.145)	(0.443)	(0.234)	(0.493)	(0.260)	(0.024)
Income (Logged	0.0723*	-0.00175	-0.0157	0.0613*	0.0340^{+}	0.0271
	(0.039)	(0.937)	(0.450)	(0.011)	(0.096)	(0.165)
Constant	2.852***	0.877***	0.849***	0.279	-0.0974	-0.319^{+}
	(0.000)	(0.000)	(0.000)	(0.156)	(0.557)	(0.053)
Observations	3421	3421	3421	3421	3421	3421

p-values in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

having parents born outside Germany—is negatively correlated with parental interest in the respondents schooling, attendance of parent's evening, meeting teachers outside appointed conferences, and being a parent representative at their school. Coefficients controlling for second generation status yield similar results adding significance to the perceived frequency of parental help with studies. This is unsurprising as language and cultural barriers may prohibit immigrant parents' ability to take an active role in their child's school life, and furthermore explains the role of immigration status in decreasing the probability of gymnasium attendance. Father's years of education overall has very little correlation with indicators of parental involvement with the exception of a decreasing likelihood of parental interest in academic performance. Mother's years of education is slightly more influential, an additional year of her education is correlated with a 1.44 percent increase in the probability of attending parent evenings and a 1.26 percent increase in the probability of being a parental representative. Surprisingly, parental socioeconomic status yields very few significant results with the exception of involvement as a parental representative. Logged income, however, yields significant positive correlations with parental interest in the respondents' education, the probability of attending a parent teacher conference, and the probability of meeting the teacher outside parent-teacher conferences.

The goal of this section was to provide benchmark OLS estimates of resource dilution by family size to compare to those of IV regression analysis. Even without correcting for family size endogeneity, OLS estimates indicate a pattern more complex than resource dilution theory purposes. While children from larger families may have less privacy in their home to learn, results indicate this does not discourage them from reading, and participating in cultural activities outside the home such as music. In regards to parental treatment and involvement, results are much more consistent with theory. Participants from larger families seem to have parents that talk with them less frequently, are less likely to consult them in decision making, ask their opinion on family matters, and show them love less frequently. Moreover, family size significantly decreases parent's interest in respondents' school life, the probability of their support with homework, and the chances of them attending parent evenings at their child's school. Important to note is that ordered logit models and logit models may be more suitable to many of these likert-scale and binary variables. A separate set of models were run using these methods in tables 4.18 through 4.20 in the , yielding results with similar patterns to linear regression and linear probability. For this reason, I apply IV regression analysis to linear models discussed in this section rather than applying instrumented family size residuals to logit and ordered logit models which are at risk of providing biased error terms.

4.2.3 IV Regression Analysis

Estimates using IV regression are reported for each resource type in tables 4.10 through 4.12 and are illustrated in figures 4.6 through 4.8 for purposes of visualization. Comparing figures 4.6-4.8 to coefficient plots of OLS estimates reveals that the introduction of multiple instruments generally depresses the magnitude of family size effects closer to zero with wider margins of error. Nevertheless, in each resource category some significant results persist.

As presented in table 4.10, environment resources remain statistically significant with the exception of frequency of reading. While family size had a modest impact on reading at a 5% significance level, instrumented family size reduces this effect close to zero with a confidence interval that lies on both sides of the y-axis. This implies that in isolation, family size has no effect on a child's frequency of reading, and that other household variables such as parental education yields a larger effect. On the other hand, the negative correlation of family size with television viewing remains consistent with OLS estimates, with an effect size outweighing years of parental education and rivaling migration status effects. Lastly, instrumented family size seems to have a stronger effect on the probability of the respondent having their own room than in OLS estimates. Namely, an additional sibling reduces the probability of a respondent having their own room by about 6 percentage points, as compared to 3.7 percentage points in OLS regression.

Stage two of instrumented effects on cultural activities are also presented in table 4.10. Consistent with OLS findings, family size effects play no significant role on participation in sports. On the other hand, while previous estimates showed positive correlation between family size and participation in music and no significant correlation in dance, IV estimates wash out significant effects of family size on music (even changing signs), and introduce significant effects on dance and performing arts. Specifically, as consistent with RDT and Blake's (1989) findings, increased family size decreases participation in dance and theater, albeit only at a 10%

Table 4.10 IV regression estimates of dilution of environmental and opportunity

	Reading	TV	Has own room	Sports	Music	Dance	Tech Work
Family Size	0.00335	**-0.104****	**-0.0654*****	0.0740	-0.1000	**-0.0835+**	**-0.0993***
	(0.954)	(0.003)	(0.000)	(0.162)	(0.142)	(0.095)	(0.037)
Female	0.422***	-0.0863***	0.00232	-0.231***	0.162**	0.575***	-0.820***
	(0.000)	(0.000)	(0.818)	(0.000)	(0.003)	(0.000)	(0.000)
Age (2024)	0.0165***	0.00546+	0.00442***	-0.0206***	-0.0304***	0.0111*	0.0209***
	(0.001)	(0.064)	(0.000)	(0.000)	(0.000)	(0.010)	(0.000)
First Generation	0.437***	-0.200+	-0.119**	0.0894	0.0965	-0.0793	-0.162
	(0.000)	(0.056)	(0.001)	(0.428)	(0.485)	(0.409)	(0.111)
Second Generation	0.0835	-0.117**	-0.159***	-0.113+	0.0688	0.0680	-0.000573
	(0.255)	(0.008)	(0.000)	(0.098)	(0.417)	(0.295)	(0.992)
Mother's Edu (yrs.)	0.0335**	-0.00943	0.00448+	0.0368***	0.0216	0.0302**	0.00125
	(0.006)	(0.184)	(0.073)	(0.000)	(0.157)	(0.006)	(0.900)
Father's Edu (yrs.)	0.0707***	-0.00726	-0.00277	0.0194+	0.0731***	0.00867	-0.0110
	(0.000)	(0.339)	(0.325)	(0.074)	(0.000)	(0.439)	(0.267)
Parental ISEI	0.00393+	-0.00153	0.00118*	0.00378+	0.000215	0.00140	0.00215
	(0.070)	(0.244)	(0.018)	(0.056)	(0.936)	(0.470)	(0.240)
Income (logged)	-0.0539	-0.0320	0.0604***	0.100	0.265**	0.189**	0.115*
	(0.445)	(0.456)	(0.000)	(0.112)	(0.001)	(0.002)	(0.039)
Constant	1.414**	5.361***	0.686***	2.493***	0.156	-0.508	1.029*
	(0.009)	(0.000)	(0.000)	(0.000)	(0.806)	(0.277)	(0.018)
Observations	3421	3421	3421	3421	3421	3421	3421

Table 4.11 IV estimates for dilution of parent treatment.

	(1) Parent Important	(2) Parent Talks	(3) Parent Consults	(4) Parent Appreciates	(5) Parent Solves	(6) Parent trusts	(7) Parent asks opinion	(8) Parent gives reason	(9) Parent shows love
Family Size	0.0112	-0.0438	-0.0554	-0.00757	0.0404	-0.0241	**-0.0996***	**-0.0867***	-0.000504
	(0.552)	(0.184)	(0.162)	(0.810)	(0.281)	(0.458)	(0.015)	(0.037)	(0.988)
Female	0.0430**	0.0846**	0.116***	0.0569*	-0.0434	-0.0434	0.0717*	-0.0301	0.0333
	(0.009)	(0.002)	(0.001)	(0.026)	(0.156)	(0.118)	(0.030)	(0.362)	(0.226)
Age (2024)	-0.00729***	-0.00530^{+}	0.00249	-0.00898**	-0.0115***	-0.00469	-0.00808*	-0.0150***	-0.0146***
	(0.000)	(0.087)	(0.495)	(0.002)	(0.001)	(0.115)	(0.034)	(0.000)	(0.000)
First Generation	0.0597	0.0332	-0.244*	-0.0873	-0.0560	0.0167	-0.150^{+}	-0.0498	0.128^{+}
	(0.146)	(0.640)	(0.011)	(0.224)	(0.486)	(0.827)	(0.096)	(0.592)	(0.079)
Second	0.0165	-0.0361	-0.0135	-0.0326	-0.0154	-0.0657	-0.120*	-0.0316	0.0646
	(0.462)	(0.395)	(0.791)	(0.420)	(0.752)	(0.129)	(0.028)	(0.542)	(0.133)
Mother's Edu	-0.0111**	0.0119^{+}	0.0208*	-0.00389	0.0138^{+}	0.00187	0.0122	0.0235**	-0.0132^{+}
	(0.010)	(0.095)	(0.016)	(0.564)	(0.087)	(0.790)	(0.161)	(0.005)	(0.070)
Father's Edu	-0.00144	0.0104	0.0153^{+}	0.00825	0.00773	-0.00685	0.00135	0.0119	0.00328
	(0.734)	(0.158)	(0.091)	(0.236)	(0.345)	(0.342)	(0.881)	(0.164)	(0.664)
Parental ISEI	-0.000645	0.000925	0.00419**	0.00318*	0.00317*	0.00137	0.00211	0.00415*	0.00155
	(0.389)	(0.481)	(0.008)	(0.011)	(0.031)	(0.313)	(0.188)	(0.011)	(0.247)
Income (logged)	0.0549*	0.127**	0.0205	0.0658	-0.00372	0.122**	0.0882	0.0909^{+}	0.115**
	(0.028)	(0.002)	(0.683)	(0.125)	(0.937)	(0.003)	(0.102)	(0.081)	(0.006)
Constant	4.615***	2.503***	2.873***	3.524***	3.496***	3.408***	2.978***	2.915***	3.749***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3421	3421	3421	3421	3421	3421	3421	3421	3421

p-values in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.12 IV estimates for dilution of parental involvement.

	(1) Parental Interest	(2) Parental Help	(3) Parent's Evening	(4) Parent teacher conference	(5) Meets Teacher	(6) Parent representative
Family Size	**-0.0601**$^{+}$	**-0.0414***	-0.0304	**-0.0436***	-0.00317	-0.0231
	(0.067)	(0.018)	(0.109)	(0.043)	(0.863)	(0.146)
Female	-0.0808***	0.00718	-0.0301*	-0.0268	-0.0493***	0.0175
	(0.001)	(0.592)	(0.032)	(0.113)	(0.001)	(0.191)
Age (2024)	0.00107	-0.00350*	-0.00205	-0.00270	0.00218	-0.00189
	(0.709)	(0.021)	(0.205)	(0.158)	(0.170)	(0.217)
First Generation	-0.147*	-0.0436	-0.147***	0.0336	-0.0860**	-0.115***
	(0.049)	(0.266)	(0.001)	(0.460)	(0.009)	(0.000)
Second Generation	-0.000727	-0.0786***	-0.0865***	-0.00861	-0.0349	-0.0742***
	(0.985)	(0.000)	(0.000)	(0.749)	(0.111)	(0.000)
Mother's Edu (yrs.)	-0.0109	0.00289	0.0143***	-0.00461	-0.00273	0.0120**
	(0.101)	(0.410)	(0.000)	(0.324)	(0.479)	(0.003)
Father's Edu (yrs.)	-0.0186**	0.00217	0.000795	-0.00642	-0.000479	0.00749^{+}
	(0.006)	(0.546)	(0.840)	(0.174)	(0.901)	(0.075)
Parental ISEI	0.00152	0.000215	0.000744	0.000129	0.000685	0.00110^{+}
	(0.196)	(0.738)	(0.265)	(0.878)	(0.313)	(0.099)
Income (logged)	0.0772*	0.00836	-0.0146	0.0773**	0.0359	0.0400^{+}
	(0.040)	(0.719)	(0.505)	(0.002)	(0.103)	(0.055)
Constant	2.859***	0.893***	0.851***	0.303	-0.0946	-0.300^{+}
	(0.000)	(0.000)	(0.000)	(0.126)	(0.568)	(0.069)
Observations	3421	3421	3421	3421	3421	3421

p-values in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

level of significance. Furthermore, the negative correlation between family size and tech work remains statically significant and strengthens in magnitude compared to OLS estimates. Examining covariates may shed light on why these changes occur. The effect size of father's years of education participation in music is strengthened by the inclusion of instrumented family size. Similarly, the effect size of mother's education on participation in dance decreases in a second stage IV regression. This indicates that the instrument is accounting for an unobserved difference between parents of different family sizes that correlates with their level of education.

IV regression analysis causes the most impactful changes on family size effects of parental treatment as seen in table 4.11. Of the five significant coefficient correlations of family size, four lose significance: correlation with frequency of talks, decision consultation, appreciation by the parent, and expression of love by the parent. The effect of family size on a parent asking the respondent an opinion on family matters remains significant and increases slightly in effect size. Covariates of this dependent variable remain consistent with the exception of first and second immigration status which decrease slightly in magnitude from OLS estimates. Family size effect on whether a parent gives a reason for a decision, a proxy for parental attention, becomes significant and negative in IV estimates. The positive correlation of mother's years of education with these variables decreases slightly with the introduction of instrumented family size.

Lastly, table 4.12 reports instrumented coefficients of parental involvement in the respondent's education. Results from this table generally agree with OLS estimates. The effects of family size on parental interest increases from OLS estimates but drops to a 10% significance level indicating that the margin of error is in between two signs. The probability of parents helping the respondent with homework decreases with family size by 4.14 percentage points using instrumented estimates versus 1.92 percentage points under OLS regression. Parallel to this, the effect size of second generation immigration status decreases in magnitude from OLS estimates. Family size effects on the probability of attending a parent's evening loses significance while effects on attending a parent teacher conference gains significance. Specifically, the birth of another child decreases the probability of attending a parent teacher conference by 4.36 percentage points. Income also has an increased effect on this variable compared to OLS estimates.

Overall, the inclusion of an instrumented measure of family size seems to complicate the story of resource dilution. Unlike predicting educational attainment outcomes, implementing multiple instruments does not consistently increase the magnitude of family size effects. In regards to environment, the inclusion of instrumented family size reveals that OLS estimates greatly exaggerated decreasing television viewing and increasing reading in larger families. Similarly, instrumental variable analysis washed out all statistical significance in regards to cultural activities with the exception of technical work. Even more devastating to RDT conjectures were IV estimates on parental treatment, which only found two indicators of parental attention to be negative and significant. The most consistent and robust finding with regards to RDT was the dilution of parental involvement. However, while some estimates of family sized strengthened marginally in magnitude, as illustrated in x, overall estimates were truncated much closer to the y-axis, indicating that in isolation, family size effects do not dilute interpersonal resources to the extent found by scholars such as Downey (1995).

However, since the instruments chosen for my analysis were found to be correlated with the error term when testing the probability of gymnasium attendance it is important to test the diagnostics of the instruments on the twenty-two dependent variables tested in this analysis. The F-statistic is consistent for all dependent variables as they implement the same controls. Just as in the previous section, first-stage IV regression yields an F-Statistic of 70, indicating that the instruments identify the endogenous variable of family size. Table 4.13 reports the Hansen-J statistic for each dependent variable used for analysis in this section. Of the twenty-two mechanisms examined, five yield p-values less than .05. This means that at least one instrument is correlated with the error terms for the following regressions: reading, sports, music, appreciation by parents, and parental attendance of parent-teacher conferences. Furthermore, three other dependent variables yield p-values that are only marginally larger than the cutoff p-value: TV viewing, having their own room, and parents consulting the respondent in decision making. As done in the previous section, I re-ran the problematic models without the instruments of father's and mother's age at first birth. This only increased the p-value in regressions with sports, music, and attendance of parent-teacher conferences as the dependent variables, implying that an unobserved factor related to parent's age may be correlated with these variables. That the p-value for the reading and parental appreciation remains small, indicates that these instruments may be problematic in other regards.

The purpose of this section was to empirically test resource dilution theory using biased OLS estimates and to see if the theory holds with instrumented family size. In chapter two, the following hypothesis was proposed based primarily off of Downey's (1995) findings:

H2: Larger sibships are correlated with the dilution of both interpersonal and cultural resources in isolation of family size endogeneity.

Results from this section cannot confirm this hypothesis. If cultural resources are considered, IV estimates show that in isolation, the impact of family size on participation in cultural activities is mostly insignificant. The primary exception to this is frequency of technical work such as computer programming, a hobby which is not traditionally considered in theory of cultural reproduction, but may nevertheless lead to more successful educational outcomes. Television viewing is also found to be significant and passes IV diagnostics, but nevertheless runs counter to theories of resource dilution, with larger families watching fewer TV. Therefore, in regards to cultural resources, evidence does not fully support RDT.

Regarding interpersonal resources, IV estimates indicate that family background characteristics play a larger role than family size on respondent's parenting styles. Nevertheless, instrumented family size seems to impact parent's propensity to help their child with homework and attend parent-teacher conferences. However, the latter finding is compromised by a violation of the exclusion restriction, indicating that for example, the mother's age may also be correlated with the probability of her attending the conference. Overall, where the instruments used in this analysis are valid, they only sporadically support resource dilution in isolation. Therefore, the second hypothesis cannot be confirmed.

4.3 Testing Confluence Theory

As previously discussed in chapter two, a formal test of confluence theory (CT) requires testing family size effects on cognitive ability. This section details the results from implementing IV analysis proposed by Jæger (2006) onto a continuous measure of cognitive ability and introducing sibling configuration variables arguably only theoretically relevant to CT.

4.2.1 Descriptive Statistics and Bivariate AnalysIs

To generate a sample appropriate for testing CT, cognitive test results from 3,142 seventeen-year old respondents are merged with biographical information of their parents found in the *bioparen* dataset. Similar to the sampling construction of the two previous sections, additional information on the respondent's parents including the instrumental variables are merged from each parent's individual questionnaires (*pl*) and the mother's birth history (*biobirth*). Additional basic information on the respondent such as gender, year of birth, and migration status are merged from their person generated datasets (*pgen* and *ppath*). Household income, as in previous studies, is also merged from the household generated dataset (*hgen*) in the year that the cognitive test was taken.

To include data on sibling configuration beyond family size, information from the generated sibling biography dataset (*biosib*) was merged into the final dataset. Using this dataset three key variables were created. First, age spacing is generated by measuring the absolute age distance between the respondent and the next sibling in the SOEP (regardless if the sibling is older or younger). Second, birth order is recorded from a continuous variable (*pos_sib*) into three dummy variables indicating if the respondent is the oldest, youngest, or a middle child. The last variable generated from this dataset is family size. A variable included in the dataset already indicates the number of siblings the respondent has included in the SOEP including the respondent herself (*num_sib*). However, since siblings in the household could be excluded from this dataset but still impact the intellectual milieu of the family in accordance with CT, I take the highest value between this variable and the number of children in the household merged from the mother's birth history. Due to the inclusion of sibling configuration variables such as age spacing and birth order, only children are necessarily excluded from this analysis, resulting in a sample size of 1,933 individuals.

Table 4.13 in the appendix reports summary statistics of the sample categorized by people who scored in the highest, middle, and lowest tertile on the test of cognition. Consistent with theory, individuals who score high on cognition come from smaller families. In fact, respondents who score in the lowest tertile have about .23 more siblings than those who score high in cognitive ability. On the other hand, respondents in the highest tertile have higher sibling density than other respondents, with respondents' next sibling on average .28 more years between them and their next sibling. This directly contradicts previous conjecture that people from families with closely-aged siblings are at a particular disadvantage since this brings down the intellectual average of the family. Birth order effects are also not as clean-cut as CT predicts. Oldest children are over-represented in both high and low tertiles, while underrepresented in the middle tertile scores. Middle children are most represented in the lowest tertile, while youngest children are found most often in the highest tertile.

While educational outcomes from the first section of this chapter found women overrepresented in the highest school track, table 4.13 finds that women are overrepresented in the lowest tertile of cognitive ability test scores. Similarly, older cohorts seem to be slightly overrepresented in the highest tertile. Respondents with no migration background are overrepresented in the highest tertile, while first and second generation migrants are overrepresented in the lowest tertile.

Parental education seems to have a large impact on cognitive ability scores. Father's of respondents in the highest tertile have about 2.3 years more of education than father's of respondents in the lowest tertile. Similarly mother's of respondents in the highest category have about 2.1 more years of education than mother's in the lowest category. The average difference between the years of education of the respondent's mother and father is lowest in the middle tertile (approximately 0.24) and equally high in the other two tertiles (0.5). In all categories, the father's number of years of education is higher.

Respondents who score highest on the cognition test also have parents who score higher on the socioeconomic scale (ISEI)—with a dramatic 6 point difference between the highest and the middle tertile, and a 12 point drop in the lowest tertile. While EGP is not included in multivariate analysis due to oversaturation, it is demonstrative evidence of class differences. Respondents with parents in EGP class I and class II are overrepresented in the highest tertile, with cases of EGP class VIIb completely unrepresented in this category. Respondents in the highest tertile are also marginally more wealthy in the highest tertile.

4.2.2 Biased OLS Estimates

Figure 4.14 reports results from three models of OLS linear regression. The first model predicts cognitive ability based on family size, including covariates of age spacing, dummy variables for birth order, gender, age, and migration status. As consistent with family size literature and CT, the birth of an additional child is correlated with a .39 score decrease in cognitive ability but only at a 10% significance level with a p-value of 0.059. With each additive model, family size not only weakens, but loses significance altogether. This implies that the covariates of each model have a stronger influence on cognitive ability than family size. Age spacing also yields a negative coefficient with a significance level of 0.05. Indicating that having a sibling further in age is correlated with lower cognitive ability. As previously mentioned, this is directly contradictory to the mechanism proposed in confluence theory. Birth order also yields significant results in this model with effect sizes that outweigh those of family size and age spacing. Accordingly, being the oldest child is correlated with a 1.104 point increase in score, whereas being the youngest is correlated with a 1.470 score increase in cognition compared to being a middle child. This finding also contradicts confluence theory, which predicts that oldest children are at an intellectual advantage due to the tutoring advantage as described in chapter three. Youngest children should only be at a great advantage intellectually if there is large spacing between them and their next older sibling, since they are born into a more intellectually mature environment.

Model one also provides estimates for gender age and migration status. Being female is correlated with 1.15 score decrease in cognitive ability with a significance level of 0.01. This correlation strengthens in magnitude in model two and weakens again with the inclusion of logged income and parental socioeconomic status, but remains statistically significant. Even though every respondent in the pooled sample took the test at 17, the age coefficient presents a slight cohort effect, whereby older cohorts score slightly higher than younger cohorts. Lastly, migration status has the strongest correlation with cognitive ability in the entire model. Being a first generation immigrant is correlated with almost a 5 point decrease in cognitive ability score. Meanwhile second generation status is correlated with a 4.7 point decrease. Both these coefficients are statistically significant at a 0.001 and persist in magnitude and significance in additive models.

The second model introduces controls for parental education. Both mother's and father's education yield highly significant results, but mother's education is slightly more influential on

Table 4.14 OLS estimates for sibling configuration on cognitive ability.

	(1) OLS1	(2) OLS2	(3) OLS3
Family Size (generated)	-0.394^{+}	-0.0839	-0.0104
	(0.059)	(0.649)	(0.954)
Age Spacing	-0.220*	-0.0861	-0.0728
	(0.020)	(0.326)	(0.406)
Oldest	1.104*	0.810	0.750
	(0.047)	(0.125)	(0.150)
Youngest	1.479*	1.149*	1.120*
	(0.014)	(0.046)	(0.050)
Female	-1.146**	-1.191**	-1.172**
	(0.004)	(0.002)	(0.002)
Age (2024)	0.120*	0.135**	0.161***
	(0.015)	(0.003)	(0.000)
First Generation	-4.972***	-3.385*	-2.960*
	(0.001)	(0.011)	(0.026)
Second Generation	-4.746***	-2.386***	-2.303***
	(0.000)	(0.000)	(0.000)
Number of Years of Education (Father)		0.605***	0.347***
		(0.000)	(0.001)
Number of Years of Education (Mother)		0.757***	0.623***
		(0.000)	(0.000)
Parental ISEI			0.0742***
			(0.000)
Income (logged)			0.758
			(0.139)
Constant	30.56***	11.55***	5.420
	(0.000)	(0.000)	(0.225)
Observations	1933	1933	1933

p-values in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

cognitive ability than that of the father. One more year of education for the respondent's mother is correlated with a 0.757 point increase in cognitive ability, while an additional year of father's education yields a 0.605 point increase. Inclusion of these variables washes out effects of family size and age spacing. However, the youngest child effect remains significant and with a fairly large effect size. Correlations of immigration status also remain negative and significant, although first generation immigration status falls down to a 0.05 level of significance. Meanwhile gender effects remain

Parental ISEI and logged income are introduced in the third model. Both yield positive coefficients but only socioeconomic status holds statistical significance—with a one unit increase in parental ISEI is correlated with a 0.074 point increase in cognitive ability. Parental Education, age, and being the youngest child remain positive and significant while immigration status and being female remain negative and significant with the inclusion of these control variables.

Overall, even OLS estimates reported in table 4.14 do not yield results consistent with previous family size literature. The primary argument raised by critics of confluence theory and sibship literature at large, is that cross sectional analysis overestimates family size effects on general cognition because it fails to account for unobserved differences between parents even in the presence of multiple control variables (Guo and Vanwey, 1999a; Rogers 2000). However, the analysis presented in table 4.14, which contains only a modest number of family-level control variables, already demonstrates that factors such as socioeconomic status, migration status, and parental education deem family size effects relatively trivial. Nevertheless, previous use of instruments such as number of siblings of each parent and parental age at first birth have demonstrated that instrumenting family size to account for its endogeneity often reveals previous underestimation of its effects on educational and cognitive outcomes. Meanwhile as discussed in previous sections of this thesis, use of these instruments may violate the assumption of exclusion restriction. For this reason, the following section applies these instruments for comparison and exploits the opportunity to test their validity in light of a new outcome variable.

4.2.3 IV Regression Analysis

Table 4.15 reports first-stage regressions for three different models. In each model both number of siblings of the respondent's mother and the respondent's father are positively correlated with the newly generated measure of family size. Predictive power of these instruments, however, is slightly weaker than the sample shown in table 4.3. Mother's age predictably has a negative correlation with family size as is consistent with previous findings. Father's age at first birth, however, has no significant correlation with family size in this sample. However, both F-statistics and Kleibergen-Paap rk LM statistics yield results for each model that indicate the model is not underidentified.

Sibling configuration indicators are also included in first-stage regressions. In each model, age spacing is negatively correlated with family size. This indicates that a larger age difference between the respondent and their next sibling is correlated with having fewer siblings. Birth order position also correlates with family size. Accordingly, with the reference category of middle child, being the youngest child is correlated with decrease in family size by 1.159 units. Being the oldest child is correlated with a much smaller decrease in family size of 1.088 units. Since all respondents in the sample necessarily have siblings these correlations follow a reasonable rationale. As expected, gender has no significant correlation with family size. Age in 2024 has yields a slight negative coefficient on family size in each model, indicating that older cohorts in the sample come from larger families. Migration status has no significant effect on family size in all three models.

Model two of first-stage regressions indicate that years of education has no significant effect on sibling size when mother's age and number of parental siblings are included in the regression. This finding persists in later models.

In model three, parental socioeconomic status and logged income are introduced into the first-stage regression. Parental ISEI is a poor predictor of family size when accounting for the model's instruments. Logged income, on the other hand, seems to have a positive effect on family size in this model. This indicates that when parents have stronger family values, having a larger income enables them to have larger families.

Table 4.15 First stage IV results for cognitive ability.

	IV1: First-Stage Regression	IV2: First-Stage Regression	IV3: First-Stage Regression
No. Of Siblings (Mother)	0.0415*	0.0471*	0.0452*
	(0.040)	(0.024)	(0.036)
No. Of Siblings (Father)	0.0422**	0.0461**	0.0438**
	(0.005)	(0.003)	(0.006)
Mother's age at first birth	-0.0610***	-0.0666***	-0.0663***
	(0.000)	(0.000)	(0.000)
Father's age at first birth	0.00444	0.00394	0.00306
	(0.537)	(0.579)	(0.660)
Age Spacing	-0.0501***	-0.0498***	-0.0483***
	(0.000)	(0.000)	(0.000)
Oldest	-1.088***	-1.082***	-1.067***
	(0.000)	(0.000)	(0.000)
Youngest	-1.159***	-1.158***	-1.151***
	(0.000)	(0.000)	(0.000)
Female	-0.0471	-0.0490	-0.0525
	(0.268)	(0.247)	(0.219)
Age (2024)	-0.0230***	-0.0240***	-0.0195**
	(0.000)	(0.000)	(0.005)
First Generation	0.359	0.371	0.347
	(0.222)	(0.204)	(0.213)
Second Generation	0.112	0.137	0.130
	(0.264)	(0.167)	(0.179)
Mother's Edu (yrs.)		0.0201	0.0188
		(0.116)	(0.151)
Father's Edu (yrs.)		0.00669	0.00810
		(0.492)	(0.449)
Parental ISEI			-0.00450
			(0.123)
Logged Income			0.194*
			(0.036)
Constant	5.846***	5.667***	4.192***
	(0.000)	(0.000)	(0.000)
Observations	1933	1933	1933

Table 4.16 Second stage IV estimates for sibling configuration on cognitive ability.

	(1) IV1	(2) IV2	(3) IV3
Family Size (generated)	-6.797***	-1.679*	-0.907
	(0.000)	(0.033)	(0.247)
Age Spacing	-0.411***	-0.140	-0.102
	(0.000)	(0.137)	(0.270)
Oldest	-7.119***	-1.205	-0.359
	(0.000)	(0.284)	(0.741)
Youngest	-6.996***	-0.926	-0.0321
	(0.000)	(0.430)	(0.978)
Female	-1.446**	-1.263**	-1.217**
	(0.004)	(0.001)	(0.001)
Age (2024)	0.0438	0.115*	0.153***
	(0.473)	(0.012)	(0.001)
First Generation	-1.433	-2.591+	-2.563+
	(0.579)	(0.089)	(0.072)
Second Generation	-2.649**	-1.986***	-2.093***
	(0.002)	(0.001)	(0.000)
Number of Years of Education (Father)		0.721***	0.610***
		(0.000)	(0.000)
Number of Years of Education (Mother)		0.573***	0.345***
		(0.000)	(0.001)
Parental ISEI			0.0661***
			(0.000)
Income (logged)			0.886
			(0.103)
Constant	58.21***	19.31***	8.783+
	(0.000)	(0.000)	(0.088)
Observations	1933	1933	1933

p-values in parentheses
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.16 reports three models with instrumented family size effects on cognitive ability. Similar to the findings of Radyakin (2007) and Jæger (2006), the introduction of the four instruments yields much stronger negative coefficients of family size. The first model reports that an additional child in the family is correlated with a dramatic 6.79 point decrease in cognitive ability at a .001 level of significance. Other sibling configuration variables also increase in magnitude and significance. Oldest children have a greater disadvantage over their younger siblings with a 7.119 point decrease in cognitive ability before accounting for parental and household level controls. Comparably, being the youngest child is correlated with a 6.996 point decrease in cognitive ability at a .001 level of significance. However, these variables lose significance with the inclusion of parental education, socioeconomic status, and logged income in later models.

Gender is also included in the first model, yielding a very similar negative significant coefficient found in OLS estimates. Age cohort effects initially have no effect on cognitive ability in IV models, but gain significance in later models---agreeing with OLS estimates that older cohorts score slightly higher than younger cohorts. Being a first generation immigrant has no significant effect in any of the second-stage IV models. Second generation status, on the other hand, is correlated with a 2.7 point decrease in test score, which remains significant and magnitude in later models.

Model 2 introduces parental education as control variables. While years of education of both respondent's parents yield positive significant results that align with OLS findings, the introduction of instrumented family size indicates that father's education has a greater impact on cognitive ability than that of the mother. This indicates that among parents who desire larger families, the father may play a larger role in social reproduction. Another notable finding in this model is that while other sibling configuration variables lose significance with the introduction of parental education controls, family size remains a significant predictor of cognitive ability. A possible interpretation of this finding is that family size may mediate social reproduction.

However, with the introduction of socioeconomic status and logged income in the third model, statistically significance for instrumented family size disappears---implying that even when family size endogeneity is accounted for, socioeconomic status plays a larger more significant role on their child's cognitive ability. In fact, IV estimates yield that a one unit increase on the ISEI scale is correlated with a .06 rise in cognitive ability. Oddly enough, even under IV estimates, however, logged income plays no significant role in cognition.

Similar to findings in the previous section, implementation of parental siblings and age at first birth as instruments seem to yield family size coefficients larger in magnitude. However, use of these instruments seems to sometimes violate the exclusion restriction, bringing into question the validity of the instruments and their corresponding estimates. For this reason, I test the null hypothesis of the Hansen-J statistic for all three models of cognitive ability. As reported in table 4.17 in the appendix, unlike estimates for upper-track attendance, the four instruments are not correlated with the error term. Paired with the high reported F-Statistic of each of these models, this gives me reason to believe that the instruments in these regressions do not yield biased estimates.

This final analysis was designed to address the third hypothesis of this thesis:

H3: Larger sibships with covariance of family structure variables is negatively correlated with cognitive ability with the application of instrumental variables.

Initial estimates of instrumented family size confirm this hypothesis. Family size remains a significant predictor of cognitive ability net of birth order and age spacing. However, estimates from the first model also indicate findings fully antithetical to confluence theory. For example, while confluence theory predicts that closely spaced siblings is additively detrimental to the effects of family size. Both OLS biased estimates and those of IV analysis indicate the opposite pattern. Furthermore, confluence theory argues that higher birth order should cushion the negative effects of family size on cognitive ability. Estimates in the current analysis once again disagree with this theory, showing that later born children are only slightly less disadvantaged than older children in cognitive ability scores. One explanation for this may be the sample composition. The current sample necessarily excludes only children. Additionally, large spacing seems to be more common in this sample. Consistent with confluence theory, younger siblings benefit from being born long after their next sibling, since they are then raised in a relatively intellectually mature environment. However, if this is true, it should also indicate that first-borns may also benefit from the tutoring effect discussed in chapter two.

Regardless of this conjecture, the introduction of robust controls in IV models immediately disregards the effects of sibling configuration. Family size effects, on the other hand, remain somewhat significant in model. As previously touched on, this could indicate that family size plays a meditative role on social reproduction. While parental education certainly has a larger impact on their child's cognitive ability, perhaps the processes of cultural reproduction are less effective in larger families. However, the introduction of parental socioeconomic status in the final model, demonstrating that even when family planning is accounted for by the instruments, ultimately socioeconomic status mitigates any possible effects of family size.

Chapter Five: Discussion and Concluding Remarks

5.1 Discussion

In chapter one, I introduced two research questions designed to be answered through novel empirical analysis. First, to what extent does family size predict the probability of educational outcomes when controlling for processes of selection bias? Results from my analysis initially agree with findings from the benchmark study conducted by Radyakin (2007) using nearly the same methods, measures, and data source. Implementing instruments of parental sibship size and age at first birth, instrumented family size coefficients indicate that the birth of an additional child is correlated with a decrease in the probability of attending the highest secondary school track (*Gymnasium*) by approximately 6.8 percentage points. Similar to previous findings, instrumented family size yields an estimate larger than biased OLS estimates by a factor of 2.5.

In other words, at first inspection, instrumenting the endogenous dependent variable by means of accounting for the parent's fertility chances and micro-culture of family values suggests that, in isolation, family size plays a highly significant role in educational attainment, rivaling that of gender, parental education, and even socioeconomic status. In fact, according to results from table 4.3, the only factors accounted for in our model in influencing the probability of attending gymnasium more strongly than family size are first generation migration status and logged income. From these findings, comes the implication that family size mitigates the process of cultural reproduction.

These results, however, are compromised by the validity of the instruments. While a simple F-Statistic bolsters evidence that number of parental siblings and each parent's age at first birth are well fitted instruments for family size, inconsistent with Jæger (2006) and Radyakin (2007), a Hansen-J statistic reveals that in the case of educational attainment, at least one instrument is correlated with the error term. How can this be interpreted? While the exclusion restriction cannot be directly tested, an inference often made from the Hansen-J statistic is that at least one instrument is correlated with the outcome variable, in this case attendance of gymnasium. To assess which of the four instruments may be problematic, I reran my models multiple times over, excluding a different instrument each time, finding that the chi-squared p-value of the Hansen-J statistic grows significantly with the exclusion of both mother's and father's age at first birth. This finding fits well into a theoretical framework. In Germany, highly-educated women are more likely to delay motherhood after they are settled into their career (Kreyenfeld and Konietyka, 2008). Older parents with more education may be better able to provide their children with the cultural capital necessary to equip them for acquiring human capital and shaping them into ripe candidates for selection into gymnasium. Moreover, Black and Deveraux (2005) argue that parental education influences the type of parent one eventually becomes. While it cannot be definitively determined whether these are the factors that account for the correlation between at least one of the instruments and the error term, what can be concluded is that the models replicated from Radyakin (2007) does not answer the research question at hand. A liberal approach to answering the question in light of these findings is excluding the probable problematic instruments. A resulting F-statistic of 27.17, while not particularly high, does exceed the rule of thumb, indicating the endogenous variable is well identified. The effect of family size in this model, while negative, is both trivial in magnitude and statically insignificant. In other words, the extent to which family size predicts educational attainment may be greatly overestimated by the studies replicated in this thesis. While other methodologies may yield different results, evidence from the current thesis suggests that family size effects on gymnasium attendance is close to null.

The second question posed in the present thesis is also considered: what mechanisms best explain the relationship between family size and educational attainment? At first glance, this conclusion seems to deem the question irrelevant. Yet I argue that empirically testing resource dilution theory (RDT) is still a worthwhile endeavor for a multitude of reasons. First, while my first analysis does not provide enough evidence to support family size effects on the probability of gymnasium attendance, there is still reasonable evidence from other studies (Lersch, 2019; Downey, 1995; Strohschein et al. 2008) that resources are necessarily scarce and compromised by the birth of additional siblings. Second, while some previous literature has tested resource dilution directly, none to my knowledge have attempted to correct these effects for family size endogeneity. Furthermore, applying these instrumental variables to models with other dependent variables provides the opportunity to test whether the instruments violate exclusion restriction is a persistent occurrence.

To test RDT, I regressed family size net of controls against twenty-one different measures of parental resources categorized by those of environment, parental treatment, and parental involvement. Benchmark results from biased OLS estimates demonstrate some evidence that agrees with RDT literature and others that directly contradict it. For example, while increasing family size predictably decreases the child's probability of having their own room, it also decreases the frequency of television viewing. Elsewhere, OLS estimates provided evidence fairly consistent with RDT, particularly in regards to parental treatment and involvement. Family size is significantly correlated with less frequent talk with parents, collaborative decision-making, encouragement or appreciation from parents, parental consultation on family matters, and even demonstration of parental love. Regarding parental involvement in child's education, OLS estimates suggest that parents with more children show less interest in their child's studies, are less likely to help with homework, and are less likely to attend parent-teacher evenings.

Implementation of instrumental variable (IV) analysis, however, yields much less promising results. IV estimates agree with OLS models that family size decreases television viewing, participation in technical work, and the probability of the respondent having their own room. Regarding parental treatment, IV analysis only results in significant negative coefficients affecting consultation in family matters and frequency of parents giving reasons for decisions made. Parental involvement also mainly loses significance with the exception of the probability that parents help with homework and attending teacher's office hours. In fact, first generation migration status seems to have a much larger effect on parental involvement in these models.

However, diagnostic tests reveal that at least one of the significant IV estimates is untrustworthy due to invalid instruments: family size effects on attendance of teacher conferences. Meanwhile other IV models that yield significance—namely frequency of television viewing and probability of the respondent having their own room—have Hansen-J p-values only marginally larger than 0.05. In light of this, results from my secondary analysis do not bolster support for RDT when accounting for selection bias.

In the final analysis, I apply the same methodological process of evaluating family size effects on cognitive ability in order to test mechanisms of confluence theory (CT). In this analysis I also introduce covariates of sibling configuration similar to Jæger (2006). Biased OLS results suggest that effects of family background, particularly migration status, outweigh and wash out any significant effects of family size. Moreover, while OLS estimates do yield significance in the most robust model for effects on birth order, the finding directly contradicts the way the mechanism is intended to function. That is, CT suggests older children should be most advantaged intellectually than youngest children, whereas my findings suggest the opposite.

Since previous studies and my primary analysis have demonstrated that OLS estimates underestimate family effects, I expected instrumenting family size based off of the parent's fertility capabilities and expectation to yield larger more significant results. IV estimates from 4.18 confirm this hypothesis, showing that in models one and two OLS underestimated family size effects on cognition by a factor of 17.25 and 20 units respectively. However, inclusion of controls in IV model three yield completely insignificant results for family size. Moreover, family configuration variables age spacing and birth order lose significance in IV model two. In short, controlling for selection bias, family size does indeed impact cognition and mitigates the effect of parental education. However, once socioeconomic status is accounted for, family size and sibling configuration are no longer relevant factors.

Unlike findings from previous analyses, IV models for cognition do not violate exclusion restriction according to their F-Statistics. However, an overarching theme in this master's thesis is the validity of using each parent's sibship size and age at first birth as instruments for educational outcomes and dilution of resources. In justifying the implementation of these instruments, Jæger (2006) argues that their empirical testability of multiple instruments rather than pure theoretical conjecture of using single instruments improved on previous attempts of adjusting for family size endogeneity.

In the present thesis, I employ the four instruments on 28 different models and find evidence that a conservative estimate of seven models violate exclusion restriction according to the Hansen-J statistic---that is, 25 % of the cases yield invalid instruments. As previously discussed, my understanding of theory in demography and social reproduction raised suspicion that parental age at first birth was the likely culprit in producing these diagnostic results. While empirical troubleshooting for gymnasium supports this theory, eliminating these instruments on models for music, reading, sports, and parental recognition does not conflate the chi-squared p-value. Furthermore, using only the two instruments seems ill-advised and less robust. As previously mentioned, the F-Statistic remains above the rule of 10 in cases of gymnasium models, indicating the use of the two instruments does not under-identify the endogenous variables. However, if the instruments are as unreliable as demonstrated in my findings, this raises speculation if methods of correcting family size endogeneity warrant further development.

5.2 Contribution, Limitations, and avenues for future research

This thesis contributes to the broader literature on family size and educational attainment by applying a relatively untouched approach to a well-studied topic. Specifically, it extends previous analyses, such as those by Radyakin (2007) and Jæger (2006), using the SOEP dataset while addressing the challenge of endogeneity in family size through instrumental variable (IV) analysis. By instrumenting family size with parental sibship size and age at first birth, this research sheds new light on the extent to which family size influences upper-track secondary school attendance (Gymnasium) and other educational outcomes.

Unlike previous studies, this thesis goes beyond simply replicating past results by testing the validity of instruments, examining mechanisms such as resource dilution (RDT) and confluence theory (CT), and exploring the robustness of the relationship between family size and educational outcomes under various model specifications. The analysis provides valuable insights into how family size may impact not only educational achievement but also parental resource allocation and cognitive abilities, thereby broadening the scope of family configuration research.

One of the central contributions is the finding that, while initial IV estimates suggest a significant negative effect of family size on Gymnasium attendance, further diagnostic tests (e.g., Hansen-J statistics) call into question the validity of certain instruments chosen by Jæger (2006) and Radyakin (2007). This reveals that, contrary to initial impressions, family size may have a much more limited effect on educational outcomes than previously thought. Additionally, this research highlights that mechanisms such as parental treatment and involvement may not consistently diminish with family size once selection bias is accounted for.

Despite these contributions, this study is not without limitations. First and foremost, the reliability of the instruments used—parental sibship size and age at first birth—proved to be problematic in several instances, as revealed by the Hansen-J statistic. While exclusion of certain instruments improved the validity of some models, the inconsistency in instrument reliability raises concerns about the robustness of IV estimates. As such, the results regarding the causal impact of family size on educational outcomes should be interpreted with caution.

Second, while this thesis tests various mechanisms (RDT and CT) that could explain the relationship between family size and educational attainment, the evidence from IV models that *do* yield validity does not strongly support resource dilution or confluence effects once selection bias and socioeconomic factors are controlled for. The complexity of these relationships, especially in a highly stratified educational system like Germany's, means that further research is needed to unpack the nuances of how family dynamics shape educational trajectories.

Additionally, the generalizability of the findings is limited. Using these instruments required sample selection that excluded children with single parents, cases of separation, or simply when information on at least one parent is missing. As indicated by other works (Miline et al., 1986, Kelly, 1998; Sun and Li 2008), this may be a huge oversight in family size literature, particularly when prenatal separation may impact the availability of parental resources. Use of these instruments also eliminates the opportunity to investigate gender composition effects. In previous work, this was examined by treating the number of brothers and the number of sisters a respondent has as discrete covariates (Powell and Steelman 1989, 1990; Kalmijn 2016). However, this is not possible when attempting to correct for family size endogeneity using IV analysis since gender composition is randomly selected.

Finally, measurement limitations in the operationalization of certain variables, such as parental involvement or the quality of resources provided to children, may not fully capture the breadth of family dynamics at play. Resources are measured by the subjective perception of the youth respondent rather than by objective measures of time and attention. Downey (1995) for example, measures parental involvement as the quantitative number of friends the parents know of the child and merges survey responses provided by the parents and the child. Since the SOEP is a household-level panel, this may be a possibility for further research.

Another avenue of research to study the mechanisms of resource allocation lies in qualitative analysis in the vein of Calarco (2011; 2014), who reveals the dynamic processes of social reproduction in the classroom. This is especially important, as past research suggests that resource allocation may be mitigated by regeneration and communal parenting. As discussed in chapter one, Germany's increasing policy emphasis on the expansion of public daycares (*Kitas*) might supplement cultural resources parents of certain class backgrounds may be unable to provide—leading to the development of RDT alterations e.g. conditional resource dilution (Gibbs et al., 2016). Qualitative studies might provide better insights into the conditions in which parental resource scarcity proves irrelevant.

Another limitation of this study is its overreliance on instrumental variables more generally. While their use is designed to adjust for selection bias, it is still susceptible to omitted variable bias, particularly with respect to time-variant family characteristics that could influence both fertility decisions and educational outcomes. An optimal research design for testing the effects of family size may be better followed in the tradition of Guo and Vanwey (1999a), by looking at dynamic longitudinal changes in family size and cognition results. While it is not possible to do this with the *cogdj* dataset, since most participants take the test only once, exploiting the longitudinal nature of the SOEP dataset may be feasible with other outcome variables.

Overall this thesis provides valuable insights into the ongoing debate on family size and educational attainment, but it also highlights the methodological challenges and complexities of identifying causal effects in this domain. While instruments initially appear to be a promising solution to the problem of selection bias and endogenous variables, future research interested in the effects of sibling configuration on social reproduction should either consider developing models with new instruments based in demographic research or look to a more longitudinal design for causal interpretations.

Bibliography

Angrist, J. and Evans, W.N., 1996. Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review* 88(3): 450–77.

Becker, B. and Reimer, D., 2010. *Vom Kindergarten bis zur Hochschule*. VS Verlag für Sozialwissenschaften.

Black, S.E., Devereux, P.J. and Salvanes, K.G., 2005. Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. *American economic review*, *95*(1), pp.437-449.

Black, S.E., Devereux, P.J. and Salvanes, K.G., 2005a. The more the merrier? The effect of family size and birth order on children's education. *The Quarterly Journal of Economics*, 120(2), pp.669-700.

Black, S.E. and Devereux, P.J., 2011. Recent developments in intergenerational mobility. *Handbook of labor economics*, *4*, pp.1487-1541.

Blake, J., 1981. Family size and the quality of children. *Demography*, 18(4), pp.421-442.

Blake, J., 1985. Number of siblings and educational mobility. *American Sociological Review*, pp.84-94.

Blake J. 1989. Family Size and Achievement. Berkeley: Univ. Calif. Press

Bourdieu, P. 1973. Cultural Reproduction and Social Reproduction. In R. Brown (Ed.), *Knowledge, Education, and Cultural Change* pp. 71-84.

Bourdieu, P. 1986. The forms of capital. In: Richardson, J., *Handbook of Theory and Research for the Sociology of Education*. pp. 241–58.

Calarco, J.M., 2011. "I need help!" Social class and children's help-seeking in elementary school. *American Sociological Review*, 76(6), pp.862-882.

Calarco, J.M., 2014. Coached for the classroom: Parents' cultural transmission and children's reproduction of educational inequalities. *American Sociological Review*, 79(5), Pp.1015-1037.

Cools, S. and Kaldager Hart, R., 2017. The effect of childhood family size on fertility in adulthood: New evidence from IV estimation. *Demography*, 54, pp.23-44.

Conley, D. and Glauber, R., 2006. 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*, 41(4), pp.722-737.

Chu, C.C., Xie, Y. and Yu, R.R., 2007. Effects of sibship structure revisited: Evidence from intrafamily resource transfer in Taiwan. *Sociology of Education*, *80*(2), pp.91-113.

Dayioğlu, M., Kirdar, M.G. and Tansel, A., 2009. Impact of sibship size, birth order and sex composition on school enrolment in urban Turkey. Oxford Bulletin of Economics and Statistics, 71(3), pp.399-426.

De Graaf, N.D. and De Graaf, P.M., 2002. Formal and popular dimensions of cultural capital: Effects on children's educational attainment.

Diaz, C.J. and Fiel, J.E., 2020. When Size Matters: IV Estimates of Sibship Size on Educational Attainment in the US. *Popul Res Policy Rev.*

DiMaggio, P. and Mohr, J., 1985. Cultural capital, educational attainment, and marital selection. *American journal of sociology*, *90*(6), pp.1231-1261.

Dodin, M., Findeisen, S., Henkel, L., Sachs, D. and Schüle, P., 2024. Social mobility in Germany. *Journal of Public Economics*, 232, p.105074.

Downey, D.B., 1995. When bigger is not better: Family size, parental resources, and children's educational performance. *American sociological review*, pp.746-761.

Downey, D.B., Powell, B., Steelman, L.C. and Pribesh, S., 1999. Much ado about siblings: Change models, sibship size, and intellectual development: Comment on Guo and VanWey. *American Sociological Review*, *64*(2), pp.193-198.

Dumais, S.A., 2006. Early childhood cultural capital, parental habitus, and teachers' perceptions. *Poetics*, *34*(2), pp.83-107.

Duta, A., Iannelli, C. and Breen, R., 2021. Social inequalities in attaining higher education in Scotland: New evidence from sibling data. *British Educational Research Journal*, 47(5), Pp.1281-1302.

Esping-Andersen, G., 1990. The three worlds of welfare capitalism. *Polity*.

Farbmacher, H., Guber, R. and Vikström, J., 2018. Increasing the credibility of the twin birth instrument. *Journal of Applied Econometrics*, *33*(3), pp.457-472.

Feng, N., 2021. The effect of sibling size on children's educational attainment: Evidence from Indonesia. *ECNU Review of Education,* 4(4), pp.830-856.

Fleury, N. and Gilles, F., 2018. The intergenerational transmission of education. A meta-regression analysis. *Education Economics*, 26(6), pp.557-573.

Freese, J. and Peterson, D., 2017. Replication in social science. *Annual Review of Sociology*, 43(1), pp.147-165.

Georg, W., 2004. Cultural capital and social inequality in the life course. *European Sociological Review*, *20*(4), pp.333-344.

Ghilagaber, G. and Wänström, L., 2015. Adjusting for selection bias in assessing the relationship between sibship size and cognitive performance. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 178(4), pp.925-944.

Gogolin, I., McMonagle, S. and Salem, T., 2019. Germany: Systemic, sociocultural and linguistic perspectives on educational inequality. *The Palgrave handbook of race and ethnic inequalities in education*, pp.557-602.

Grätz, M., 2015. When growing up without a parent does not hurt: Parental separation and the compensatory effect of social origin. *European Sociological Review*, 31(5), pp.546-557.

Grätz, M. and Torche, F., 2016. Compensation or reinforcement? The stratification of parental responses to children's early ability. *Demography*, 53(6), pp.1883-1904.

Grätz, M., 2018. Competition in the family: inequality between siblings and the intergenerational transmission of educational advantage. *Sociological Science*, *5*, pp.246-269.

Guo, G. and VanWey, L.K., 1999. Sibship size and intellectual development: Is the relationship causal?. *American Sociological Review*, 64(2), pp.169-187.

Guo, G. and VanWey, L.K., 1999a. The effects of closely spaced and widely spaced sibship size on intellectual development. *American Sociological Review*, 64(2), pp.199-206.

Hanif, M., 2014. *The relationship between sibship size and educational achievement*.

Georgetown University.

Härkönen, J., 2014. Birth order effects on educational attainment and educational transitions in West Germany. *European sociological review*, *30*(2), pp.166-179.

Henninges, M., Traini, C. and Kleinert, C., 2019. Tracking and sorting in the German educational system. *Leibniz Institute for Educational Trajectories (LIfBi) Working Paper No*, *83*.

Jacob, M., 2011. Do brothers affect their sisters' chances to graduate? An analysis of sibling sex composition effects on graduation from a university or a Fachhochschule in Germany. *Higher Education*, 61, pp.277-291.

Jæger, M.M., 2006. Does Sibship Size Affect Educational Attainment? A Reevaluation of the Resource Dilution Hypothesis Using Instrumental Variables. The Danish National Institute of Social Research, Social policy and welfare working paper #05:2006.

Jæger, M.M., 2009. Sibship size and educational attainment. A joint test of the confluence model and the resource dilution hypothesis. *Research in social stratification and mobility*, 27(1), Pp.1-12.

Jæger, M.M. and Breen, R., 2016. A dynamic model of cultural reproduction. *American Journal of Sociology*, *121*(4), pp.1079-1115.

Jonsson, J.O. and Erikson, R., 2000. Understanding educational inequality: The Swedish experience. *L'Année sociologique* (1940/1948-), pp.345-382.

Kalmijn, M. and van de Werfhorst, H.G., 2016. Sibship size and gendered resource dilution in different societal contexts. *PloS one*, *11*(8), p.e0160953.

Kallunki, J. and Purhonen, S., 2017. Intergenerational transmission of cultural capital in Finland. *Research on Finnish society*, *10*(1), pp.101-111.

Kelly, J.B., 1998. Marital conflict, divorce, and children's adjustment. *Child and Adolescent Psychiatric Clinics*, *7*(2), pp.259-271.

Klein, M., Barg, K. and Kühhirt, M., 2019. Inequality of educational opportunity in East and West Germany: Convergence or continued differences?. *Sociological Science*, 6, pp.1-26.

Kraaykamp, G. and Van Eijck, K., 2010. The intergenerational reproduction of cultural capital: A threefold perspective. *Social forces*, *89*(1), pp.209-231.

Kravdal, Ø., 2001. The high fertility of college educated women in Norway: An artefact of the separate modelling of each parity transition. *Demographic research*, *5*, pp.187-216.

Kreyenfeld, M., 2002. Time-squeeze, partner effect or self-selection? An investigation into the positive effect of women's education on second birth risks in West Germany. *Demographic research*, *7*, pp.15-48.

Kreyenfeld, M. and Konietzka, D., 2008. Education and fertility in Germany. In *Demographic change in Germany: The economic and fiscal consequences* (pp. 165-187). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kuo, H.H.D. and Hauser, R.M., 1997. How does size of sibship matter? Family configuration and family effects on educational attainment. *Social Science Research*, *26*(1), pp.69-94.

Lersch, P.M., 2019. Fewer siblings, more wealth? Sibship size and wealth attainment. *European Journal of Population*, *35*(5), pp.959-986.

Malik, A.H. and Mohamed, A.E.A., 2014. English as cultural capital: EFL teachers' perceptions: A cross-cultural study. *structure*, *5*(2), pp.10-5296.

Milne, A.M., Myers, D.E., Rosenthal, A.S. and Ginsburg, A., 1986. Single parents, working mothers, and the educational achievement of school children. *Sociology of education*, pp.125-139.

Nilsen, S.A., Breivik, K., Wold, B., Askeland, K.G., Sivertsen, B., Hysing, M. and Bøe, T., 2020. Divorce and adolescent academic achievement: Heterogeneity in the associations by parental education. *PLoS One*, *15*(3), p.e0229183.

Öberg, S., 2017a. An introduction to using twin births as instrumental variables for sibship size. Göteborg Papers in Economic History No. 22. Gothenburg: Department of Economy and Society, School of Business, Economics and Law.

Öberg, S., 2017b. Too many is not enough: Studying how children are affected by their number of siblings and resource dilution in families. *The History of the Family*, *22*(2-3), pp.157-174.

Öberg, S. 2019. A Critical Introduction to Instrumental Variables for Sibship Size Based on

Twin Births. *Studies in economic history*, pp.55–82. doi:https://doi.org/10.1007/978-3-319-99480-2_4.

Öberg, S., 2021. The casual effect of fertility: The multiple problems with instrumental variables for the number of children in families.

Ostner, I., 2010. Farewell to the Family as We Know it: Family Policy Change in Germany. *German Policy Studies/Politikfeldanalyse*, 6(1).

Page, E.B. and Grandon, G.M., 1979. Family configuration and mental ability: Two theories contrasted with US data. *American Educational Research Journal*, *16*(3), pp.257-272.

Park, H., 2008. Public policy and the effect of sibship size on educational achievement: A comparative study of 20 countries. *Social Science Research*, 37(3), pp.874-887.

Parr, N., 2006. Do children from small families do better?. *Journal of Population Research*, 23, Pp.1-25.

Peter, T., Edgerton, J.D. and Roberts, L.W., 2010. Welfare regimes and educational inequality: a cross‐national exploration. *International studies in Sociology of Education*, *20*(3), pp.241-264.

Pietsch, M. and Stubbe, T.C., 2007. Inequality in the transition from primary to secondary school: School choices and educational disparities in Germany. *European Educational Research Journal*, *6*(4), pp.424-445.

Powell, B. and Steelman, L.C., 1989. The liability of having brothers: Paying for college and the sex composition of the family. *Sociology of Education*, pp.134-147.

Powell, B. and Steelman, L.C., 1990. Beyond sibship size: Sibling density, sex composition, and educational outcomes. *Social Forces*, *69*(1), pp.181-206.

Radyakin, S., 2007. Instrumental variable analysis of the quantity-quality trade-off in German children's educational achievement. *Institute of Labor Economics Working Paper*, Leibniz University of Hannover.

Richter, D., Rohrer, J., Metzing, M., Nestler, W., Weinhardt, M. and Schupp, J., 2017. *SOEP scales manual (updated for SOEP-Core v32. 1)* (No. 423). SOEP survey papers.

Rodgers, J.L., 1984. Confluence effects: Not here, not now. *Developmental Psychology*, 20(2), p.321.

Rodgers, J.L., Cleveland, H.H., Van Den Oord, E. and Rowe, D.C., 2000. Resolving the debate over birth order, family size, and intelligence. *American Psychologist*, 55(6), p.599.

Schneider, T., 2008. Social inequality in educational participation in the German school system in a longitudinal perspective: Pathways into and out of the most prestigious school track. European Sociological Review, 24(4), pp.511-526.

Shavit, Y. and Pierce, J.L., 1991. Sibship size and educational attainment in nuclear and extended families: Arabs and Jews in Israel. *American sociological review*, pp.321-330.

Sheppard, P. and Monden, C., 2020. When does family size matter? Sibship size, socioeconomic status and education in England. Evolutionary Human Sciences, 2, p.e51.

Staiger, D.O. and Stock, J.H., 1994. Instrumental variables regression with weak instruments.

Steelman, L.C. and Mercy, J.A., 1980. Unconfounding the confluence model: A test of sibship size and birth-order effects on intelligence. *American Sociological Review*, pp.571-582.

Steelman, L.C., 1985. A tale of two variables: A review of the intellectual consequences of sibship size and birth order. *Review of Educational Research*, *55*(3), pp.353-386.

Steelman, L.C., Powell, B., Werum, R. and Carter, S., 2002. Reconsidering the effects of sibling configuration: Recent advances and challenges. *Annual review of sociology*, *28*(1), pp.243-269.

Strohschein, L., Gauthier, A.H., Campbell, R. and Kleparchuk, C., 2008. Parenting as a dynamic process: A test of the resource dilution hypothesis. *Journal of Marriage and Family*, 70(3), pp.670-683.

Sullivan, A., 2001. Cultural capital and educational attainment. *Sociology*, *35*(4), pp.893-912.

Sun, Y. and Li, Y., 2008. Parents' marital disruption and its uneven effect on children's academic performance—A simulation model. *Social science research*, *37*(2), pp.449-460.

Sun, Y. and Li, Y., 2009. Parental divorce, sibship size, family resources, and children's academic performance. *Social science research*, *38*(3), pp.622-634.

Teachman, J.D., 1987. Family background, educational resources, and educational attainment. *American sociological review*, pp.548-557.

Tramonte, L. and Willms, J.D., 2010. Cultural capital and its effects on education outcomes. *Economics of Education Review*, *29*(2), pp.200-213.

Uvaag, S.A., 2023. Siblings' educational mobility and the educational stratification of families. *British Journal of Sociology of Education*, 44(5), pp.824-842.

Velandia, W., Grandon, G.M. and Page, E.B., 1978. Family size, birth order, and intelligence in a large South American sample. *American Educational Research Journal*, 15(3), pp.399-416.

Van Eijck, K., 1997. The impact of family background and educational attainment on cultural consumption: A sibling analysis. *Poetics*, *25*(4), pp.195-224.

West, A. and Nikolai, R., 2013. Welfare regimes and education regimes: Equality of opportunity and expenditure in the EU (and US). *Journal of social policy*, *42*(3), pp.469-493.

Wikle, J.S., Jensen, A.C. and Hoagland, A.M., 2018. Adolescent caretaking of younger siblings. *Social Science Research*, *71*, pp.72-84.

Wolfinger, N.H., Kowaleski‐Jones, L. and Smith, K.R., 2003. Double impact: What sibling data can tell us about the long‐term negative effects of parental divorce. *Social Biology*, *50*(1-2), pp.58-76.

Xu, J., 2008. Sibship size and educational achievement: The role of welfare regimes cross-nationally. *Comparative Education Review*, 52(3), pp.413-436.

Zajonc, R.B. and Markus, G.B., 1975. Birth order and intellectual development. *Psychological Jereview*, 82(1), p.74.

Zajonc, R.B., 1983. Validating the confluence model. *Psychological Bulletin*, *93*(3), p.457.

Zhelenkova, A. and Panichella, N., 2023. Family socioeconomic status and sibling correlations in upper secondary education. An empirical analysis of educational inequalities in Italy. *The British Journal of Sociology*, *74*(5), pp.808-816.

Appendix

Table 2.1 Types of resources in RDT and their indicators.

Resources	**Measures**
Environment	Income, Wealth, Privacy, Educational materials
Opportunities	Income, Wealth, Cultural activities
Treatment	Time, Attention, Involvement

Table 3.1 Dependent variables and their corresponding indicators.

Dependent Variables	**Indicators**
Educational Attainment	*jl0125_v3, ks_gen_h*
Cognitive Ability	*sdindex*
Resources	(See Mechanisms of RDT)

Table 3.2 Independent control variables and their corresponding indicators.

Independent Variables	**Indicators**
Sibship size	*nums, numb*
Sibling Density	*gebsib1…gebsib11*
Birth Order	*pos_sib*
Parental Socioeconomic Status	*fisei88, misei88*
Parental Education	*fsedu, msedu, d11109*
Immigration Status	*migback*
Household Income	*hghinc (transformed into logged units)*
Gender	*sex*

Table 3.3 Instrumental variables and their indicators

Instrumental Variables	**Indicators**
Number of Siblings - Mother	*pld0030, pld0032*
Number of Siblings - Father	*pld0030, pld0032*
Age at first birth - Mother	*mybirth, kidgeb01*
Age at first birth - Father	*fybirth, kidgeb01*

Table 3.4 Mechanisms of RDT used as dependent variables in second analysis.

Mechanisms of RDT	**Variables**
Environment/Setting	● How often do you read (jl0066) ● How often do you watch TV (jl0058) ● Has own room (jl0006)
Opportunities	● *Cultural opportunities*: ○ How often do you play sports (jl0063) ○ How often do you do technical work or programing (jl0071) ○ How often do you play music or sing (jl0065) ○ How often do you dance or act (jl0064)
Treatment	● *Relationship*: ○ Mother/Father important in life (*jl0027/jl0026*) ○ Mother/Father talks about things you do (*jl0040/jl0041*) ○ Mother/Father asks before making decisions *(jl0044/jl0045*) ○ Mother/Father asks opinion on something you do (*jl0046/jl0047*) ○ Mother/Father able to solve problems with you (jl0048/jl0049)

	○ Mother/Father has impression of trusting you (jl0050/jl0051) ○ Mother/Father asks your decision on family matter (jl0052/jl0053) ○ Mother gives reason for making decision (jl0054/jl0055) ○ Mother/Father show they loves you (jl0056/jl0057) ● *Parental Involvement*: ○ Parents show interest in performance (jl0168) ○ Parents help with studying (jl0169) ○ Parents take part in parent's evening (jl0171) ○ Parents come to teacher's office hours (jl0172) ○ Parents visit teacher outside office hours (jl0173) ○ Parents involved as a representative (jl0174)

Table 4.1 Summary Statistics by attendance of school track.

Independent Variable	**Gymnasium**	**Other Tracks**	**Total**
Family Size	**2.647592** .0252867	**2.868716** .030842	**2.766387** .0203607
Female	**.5246459** .0118903	**.4802343** .0110399	**.5007866** .0080972
Age in 2024	**31.31785** .1249086	**32.9019** .1212484	**32.16885** .0880109
No Migration Background	**.8379603** .0087735	**.7232796** .0098857	**.7763503** .0067481

First Generation	**.0379603** .00455	**.056613** .0051067	**.0479811** .0034612
Second Generation	**.1240793** .0078493	**.2201074** .0091552	**.1756686** .0061626
Father's Education (years)	**14.00822** .0710897	**11.58638** .0498492	**12.70713** .0467056
Mother's Education (years)	**13.5119** .0648971	**11.38897** .0477052	**12.37139** .0430372
Father has Abitur	**.3988669** .0116587	**.1059053** .0067996	**.2414788** .0069309
Mother has Abitur	**.3665722** .0114731	**.0922401** .0063941	**.2191924** .0066996
Highest Parental ISEI	**58.09518** .36611995	**44.78721** .268025	**50.94573** .0034969
(I) Higher Managerial and Professional Workers"	**.372238** .0115096	**.1303075** .0074388	**.2422653** .0069386
(II) Lower Managerial and Professional Workers	**.3286119** .0111835	**.2166911** .0091038	**.2684845** .0071769
(IIIa) Routine Clerical Workers	**.1303116** .0080154	**.1410444** .0076913	**.1360776** .0055526
(IIIb) Routine Service and Sales Work	**.0844193** .0066194	**.1625183** .0081522	**.1263765** .005381
(IVa) Small Self-Employed with Employees	**.0062323** .0018738	**.0151293** .0026973	**.0110121** .00169
(IVb) Small Self-Employed without Employees	**.0101983** .0023922	**.0209858** .0031673	**.0159937** .0020316
(VI) Skilled Manual Workers	**.0390935** .0046147	**.1761835** .0084185	**.1127425** .0051219
(VIIa) Semi- and Unskilled Manual Workers	**.0266289** .0038332	**.124939** .0073064	**.0794442** .0043795

(VIIb) Agricultural Labour	**.0016997** .0009808	**.0087848** .002062	**.005506** .0011984
(IVc) Self-Employed Farmers	**.0005666** .0005666	**.0034163** .0012893	**.0020975** .0007409
Income Logged	**8.280098** .0108392	**7.961297** .0094738	**8.108828** .0075946
Observations (n)	1765	2049	3814

Mean is reported in bold font with standard deviation reported underneath.

Table 4.2 OLS regression estimates for Gymnasium attendance.

	(1) OLS1	(2) OLS2	(3) OLS3
Family Size	-0.0359***	-0.0273***	-0.0274***
	(0.000)	(0.000)	(0.000)
Female	0.0536***	0.0664***	0.0678***
	(0.001)	(0.000)	(0.000)
Age (2024)	-0.0144***	-0.0105***	-0.00863***
	(0.000)	(0.000)	(0.000)
First Generation	-0.0615	0.0723*	0.107**
	(0.103)	(0.039)	(0.002)
Second Generation	-0.157***	0.0103	0.0121
	(0.000)	(0.649)	(0.586)
Number of Years of Education (Father)		0.0481***	0.0298***
		(0.000)	(0.000)
Number of Years of Education (Mother)		0.0407***	0.0320***
		(0.000)	(0.000)
Parental ISEI			0.00426***
			(0.000)
Income (Logged)			0.0876***
			(0.000)
Constant	1.028***	-0.276***	-0.927***
	(0.000)	(0.000)	(0.000)
Observations	3814	3814	3814

p-values in parentheses
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.3 First and second stage IV estimates for Gymnasium attendance.

	(1)	(2)	(3)	(4)	(3)	(3)
	IV1: First-stage regression	IV1: Second-stage regression	IV2: First-stage regression	IV2: Second-stage regression	IV3: First-stage regression	IV3: Second-stage regression
Family Size		-0.310***		-0.0925***		-0.0680***
		(0.000)		(0.000)		(0.000)
No. of Siblings (Mother)	0.0750***		0.0853***		0.0836***	
	(0.000)		(0.000)		(0.000)	
No. of Siblings (Father)	0.0701***		0.0810***		0.0799***	
	(0.000)		(0.000)		(0.000)	
Mother's Age (first birth)	-0.0783***		-0.0904***		-0.0916***	
	(0.000)		(0.000)		(0.000)	
Father's Age (first birth)	0.00469		0.00350		0.00262	
	(0.523)		(0.628)		(0.701)	
Female	0.0456	0.0699***	0.0531	0.0701***	0.0441	0.0696***
	(0.204)	(0.000)	(0.137)	(0.000)	(0.212)	(0.000)
Age (2024)	-0.0550***	-0.0242***	-0.0558***	-0.0129***	-0.0448***	-0.00969***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
First Generation	0.423+	0.146*	0.474*	0.119**	0.478*	0.135***
	(0.055)	(0.048)	(0.032)	(0.002)	(0.023)	(0.000)
Second Generation	0.203+	-0.00994	0.275*	0.0415	0.250*	0.0301
	(0.072)	(0.811)	(0.017)	(0.103)	(0.025)	(0.210)
Father's Edu (yrs)			0.0402***	0.0479***	0.0229*	0.0298***
			(0.000)	(0.000)	(0.041)	(0.000)
Mother's Edu (yrs)			0.0250*	0.0391***	0.0181	0.0311***
			(0.050)	(0.000)	(0.167)	(0.000)
Parental ISEI					-0.00500+	0.00380***
					(0.062)	(0.000)
Income (logged)					0.498***	0.107***
					(0.000)	(0.000)
Constant	5.967***	2.059***	5.448***	-0.00558	1.682*	-0.905***
	(0.000)	(0.000)	(0.000)	(0.958)	(0.018)	(0.000)
Observations	3814	3814	3814	3814	3814	3814

p-values in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.4 Mean response regarding home setting and cultural activities by family size.

Resource	One Child	Two children	3-4 children	5+ children	Total
Reading	**3.261589** .0762373	**3.229894** .038742	**3.329299** .0339112	**3.147186** .0933509	**3.272727** .0234624
TV	**4.798013** .0305211	**4.698027** .018651	**4.605096** .0192536	**4.160173** .0831202	**4.627887** .0131894
Own Room (1 = Yes)	**.9834437** .0073548	**.9385432** .0066179	**.8853503** .0080433	**.7489177** .0285931	**.9052909** .005007
Sport	**3.622517** .0690145	**3.585736** .0336185	**3.659873** .0307971	**3.601732** .0824707	**3.624087** .0208713
Music	**2.059603** .0850104	**2.366464** .0439442	**2.440764** .0404123	**2.376623** .103962	**2.37416** .0271797
Dance	**2.036424** .073614	**2.079666** .0342704	**1.93121** .0296295	**1.995671** .0828846	**2.002046** .0208271
Tech-Work	**2.162252** .0761446	**1.939302** .0332162	**1.887261** .0304569	**1.701299** .0689233	**1.91903** .0206945
Total (*n*)	302	1318	1570	231	3421

Variable coded by frequency of participation, from never (1) to daily (5). Own room is coded as a dummy variable.

Table 4.5 Mean response regarding parental attention and treatment by family size.

Resource	One Child	Two children	3-4 children	5+ children	Total
Parent is important in your life	**4.725166** .0270857	**4.689302** .0134073	**4.702229** .0120055	**4.681818** .0310927	**4.697895** .0081939
Parent talks	**3.635762** .0444907	**3.656677** .0227754	**3.584395** .0206956	**3.452381** .0506826	**3.607863** .013965
Parent consults you before making decisions	**3.90894** .0512485	**3.698407** .028911	**3.647134** .0246849	**3.441558** .0653591	**3.676118** .0171686
Parents express appreciation for things you do	**4.092715** .0421377	**4.001138** .0218439	**3.978025** .019294	**3.885281** .0468688	**3.990792** .0131705
Parents able to solve problems with you	**3.582781** .0532161	**3.631639** .0252469	**3.624204** .02261	**3.510823** .0569345	**3.615756** .0154663
Parents have impression of trusting you	**4.278146** .0440187	**4.162367** .022532	**4.141401** .0204791	**4.084416** .0520796	**4.157702** .0138388
Parents asks opinion on family matter	**3.620861** .0551358	**3.524659** .027091	**3.384076** .0252128	**3.214286** .0579048	**3.447676** .016871
Parents give reason for making decision	**3.619205** .055073	**3.614188** .0266041	**3.542675** .0245044	**3.441558** .0613483	**3.570155** .0165165
Parents show they love you	**4.342715** .0417936	**4.197269** .0226589	**4.192994** .0206608	**4.155844** .0502936	**4.205349** .0138445
Total (*n*)	302	1318	1570	231	3421

Variables coded by frequency of occurrence, from never (1) to very often (5).

Table 4.6 Mean response of parental involvement by family size.

Resource	One Child	Two children	3-4 children	5+ children	Total
Parents show interest in performance (1= None, 4= Very Strong)	**3.082781** .040204	**3.06525** .0191474	**2.97707** .0186975	**2.865801** .0496249	**3.012862** .0123602
Parents help with studying (Yes/No)	**.807947** .0227049	**.823217** .010512	**.7898089** .0102862	**.6883117** .0305414	**.7974277** .0068726
Parents take part in parent's evening (Yes/No)	**.7980132** .0231411	**.8179059** .0106342	**.7471338** .0109732	**.6406926** .0316369	**.7717042** .0071773
Parents come to teacher's office hours (Yes/No)	**.6125828** .0280795	**.5834598** .0135844	**.5694268** .0125006	**.5844156** .0324958	**.5796551** .0084406
Parents visit teacher outside office hours (Yes/No)	**.2251656** .0240753	**.2056146** .0111365	**.2286624** .0106025	**.1991342** .0263323	**.2174803** .0070541
Parents involved as a representative (Yes/No)	**.1854305** .0224012	**.1798179** .0105823	**.2152866** .0103765	**.1125541** .0208395	**.1920491** .0067357
Total (*n*)	302	1318	1570	231	3421

Table 4.7 OLS estimates of dilution of environment and cultural opportunities.

	(1) Reading	(2) TV	(3) Has own room	(4) Sport	(5) Music	(6) Dance	(7) Tech Work
Family Size	**0.0441***	**-0.112*****	**-0.0373*****	0.00823	0.0705**	-0.0225	**-0.0514*****
	(0.026)	(0.000)	(0.000)	(0.639)	(0.004)	(0.219)	(0.001)
Female	0.421***	-0.0860***	0.00135	-0.228***	0.156**	0.573***	-0.822***
	(0.000)	(0.001)	(0.893)	(0.000)	(0.004)	(0.000)	(0.000)
Age (2024)	0.0177***	0.00522^{+}	-0.00358**	-0.0225***	-0.0253***	0.0129**	0.0223***
	(0.000)	(0.072)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
First Generation	0.408***	-0.194^{+}	-0.139***	0.137	-0.0260	-0.123	-0.196*
	(0.000)	(0.051)	(0.000)	(0.209)	(0.836)	(0.177)	(0.040)
Second Generation	0.0674	-0.114**	-0.170***	-0.0871	0.00156	0.0439	-0.0194
	(0.336)	(0.007)	(0.000)	(0.185)	(0.985)	(0.482)	(0.730)
Father's Edu (yrs)	0.0705***	-0.00721	-0.00294	0.0198^{+}	0.0721***	0.00829	-0.0113
	(0.000)	(0.345)	(0.294)	(0.068)	(0.000)	(0.462)	(0.253)
Mother's Edu (yrs)	0.0344**	-0.00960	0.00508*	0.0354***	0.0252^{+}	0.0315**	0.00227
	(0.005)	(0.174)	(0.041)	(0.001)	(0.094)	(0.005)	(0.818)
Parental ISEI	0.00442*	-0.00163	0.00152**	0.00300	0.00225	0.00213	0.00272
	(0.034)	(0.195)	(0.002)	(0.118)	(0.360)	(0.255)	(0.114)
Income (logged)	-0.0724	-0.0282	0.0477***	0.130*	0.188*	0.161**	0.0930^{+}
	(0.271)	(0.465)	(0.000)	(0.025)	(0.012)	(0.004)	(0.074)
Constant	1.386*	5.367***	0.667***	2.538***	0.0379	-0.551	0.996*
	(0.011)	(0.000)	(0.000)	(0.000)	(0.951)	(0.232)	(0.022)
Observations	3421	3421	3421	3421	3421	3421	3421

p-values in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.8 OLS estimates for dilution of parental treatment.

	(1) Parent Important	(2) Parent Talks	(3) Parent Consults	(4) Parent Appreciates	(5) Parent solves	(6) Parent trusts	(7) Parent asks opinion	(8) Parent gives reason	(9) Parent shows love
Family Size	-0.0109	-0.0308*	-0.0480***	-0.0297*	-0.00676	-0.0216^{+}	-0.0732***	-0.0227	-0.0350*
	(0.121)	(0.013)	(0.001)	(0.015)	(0.617)	(0.085)	(0.000)	(0.166)	(0.015)
Female	0.0438**	0.0841**	0.116***	0.0577*	-0.0417	-0.0434	0.0708*	-0.0323	0.0345
	(0.008)	(0.003)	(0.001)	(0.024)	(0.171)	(0.117)	(0.032)	(0.327)	(0.209)
Age (2024)	-0.00795***	-0.00491^{+}	0.00271	-0.00964***	-0.0129***	-0.00462	-0.00729*	-0.0131***	-0.0157***
	(0.000)	(0.096)	(0.437)	(0.000)	(0.000)	(0.106)	(0.044)	(0.000)	(0.000)
First Generation	0.0756^{+}	0.0238	-0.249**	-0.0714	-0.0221	0.0149	-0.169*	-0.0957	0.153*
	(0.054)	(0.724)	(0.007)	(0.305)	(0.774)	(0.838)	(0.049)	(0.260)	(0.029)
Second Generation	0.0252	-0.0412	-0.0164	-0.0239	0.00317	-0.0667	-0.130*	-0.0569	0.0782^{+}
	(0.232)	(0.315)	(0.740)	(0.537)	(0.946)	(0.109)	(0.015)	(0.246)	(0.060)
Father's Edu (yrs.)	-0.00131	0.0103	0.0152^{+}	0.00838	0.00802	-0.00687	0.00118	0.0115	0.00349
	(0.760)	(0.162)	(0.093)	(0.229)	(0.327)	(0.342)	(0.896)	(0.180)	(0.643)
Mother's Edu (yrs.)	-0.0116**	0.0121^{+}	0.0210*	-0.00437	0.0128	0.00193	0.0128	0.0249**	-0.0139^{+}
	(0.007)	(0.087)	(0.015)	(0.516)	(0.111)	(0.783)	(0.141)	(0.003)	(0.055)
Parental ISEI	-0.000909	0.00108	0.00428**	0.00292*	0.00260^{+}	0.00140	0.00243	0.00491**	0.00114
	(0.212)	(0.384)	(0.005)	(0.016)	(0.068)	(0.282)	(0.114)	(0.002)	(0.371)
Income (logged)	0.0649**	0.121**	0.0171	0.0758^{+}	0.0177	0.121**	0.0762	0.0618	0.130**
	(0.007)	(0.002)	(0.720)	(0.068)	(0.690)	(0.002)	(0.134)	(0.205)	(0.001)
Constant	4.631***	2.494***	2.868***	3.540***	3.528***	3.406***	2.959***	2.871***	3.773***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3421	3421	3421	3421	3421	3421	3421	3421	3421

p-values in parentheses $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.9 OLS estimates of dilution of parental involvement.

	(1) Parental Interest	(2) Parental Help	(3) Parent's Evening	(4) Parent teacher Conference	(5) Meets Teacher	(6) Parent Representative
Family Size	-0.0493***	-0.0192**	-0.0280***	-0.00824	0.000843	0.00525
	(0.000)	(0.002)	(0.000)	(0.292)	(0.889)	(0.379)
Female	0.0811***	0.00641	-0.0302*	-0.0280^{+}	0.0495***	0.0166
	(0.001)	(0.633)	(0.032)	(0.097)	(0.001)	(0.216)
Age (2024)	0.00139	-0.00283*	-0.00198	-0.00164	0.00230	-0.00105
	(0.607)	(0.043)	(0.192)	(0.362)	(0.122)	(0.471)
First Generation	-0.155*	-0.0596	-0.149***	0.00818	-0.0889**	-0.136***
	(0.032)	(0.106)	(0.001)	(0.845)	(0.004)	(0.000)
Second Generation	-0.00498	-0.0873***	-0.0874***	-0.0225	-0.0365^{+}	-0.0854***
	(0.894)	(0.000)	(0.000)	(0.378)	(0.081)	(0.000)
Father's Edu (yrs.)	-0.0187**	0.00203	0.000780	-0.00664	-0.000504	0.00731^{+}
	(0.006)	(0.573)	(0.843)	(0.159)	(0.895)	(0.081)
Mother's Edu (yrs.)	-0.0106	0.00337	0.0144***	-0.00386	-0.00265	0.0126**
	(0.109)	(0.333)	(0.000)	(0.405)	(0.488)	(0.002)
Parental ISEI	0.00165	0.000480	0.000773	0.000551	0.000732	0.00144*
	(0.145)	(0.443)	(0.234)	(0.493)	(0.260)	(0.024)
Income (Logged	0.0723*	-0.00175	-0.0157	0.0613*	0.0340^{+}	0.0271
	(0.039)	(0.937)	(0.450)	(0.011)	(0.096)	(0.165)
Constant	2.852***	0.877***	0.849***	0.279	-0.0974	-0.319^{+}
	(0.000)	(0.000)	(0.000)	(0.156)	(0.557)	(0.053)
Observations	3421	3421	3421	3421	3421	3421

p-values in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.10 IV regression estimates of dilution of environment and opportunity.

	Reading	TV	Has own room	Sports	Music	Dance	Tech Work
Family Size	0.00335	**-0.104****	**-0.0654*****	0.0740	-0.1000	**-0.0835**+	**-0.0993***
	(0.954)	(0.003)	(0.000)	(0.162)	(0.142)	(0.095)	(0.037)
Female	0.422***	-0.0863***	0.00232	-0.231***	0.162**	0.575***	-0.820***
	(0.000)	(0.000)	(0.818)	(0.000)	(0.003)	(0.000)	(0.000)
Age (2024)	0.0165***	0.00546+	0.00442***	-0.0206***	-0.0304***	0.0111*	0.0209***
	(0.001)	(0.064)	(0.000)	(0.000)	(0.000)	(0.010)	(0.000)
First Generation	0.437***	-0.200+	-0.119**	0.0894	0.0965	-0.0793	-0.162
	(0.000)	(0.056)	(0.001)	(0.428)	(0.485)	(0.409)	(0.111)
Second Generation	0.0835	-0.117**	-0.159***	-0.113+	0.0688	0.0680	-0.000573
	(0.255)	(0.008)	(0.000)	(0.098)	(0.417)	(0.295)	(0.992)
Mother's Edu (yrs.)	0.0335**	-0.00943	0.00448+	0.0368***	0.0216	0.0302**	0.00125
	(0.006)	(0.184)	(0.073)	(0.000)	(0.157)	(0.006)	(0.900)
Father's Edu (yrs.)	0.0707***	-0.00726	-0.00277	0.0194+	0.0731***	0.00867	-0.0110
	(0.000)	(0.339)	(0.325)	(0.074)	(0.000)	(0.439)	(0.267)
Parental ISEI	0.00393+	-0.00153	0.00118*	0.00378+	0.000215	0.00140	0.00215
	(0.070)	(0.244)	(0.018)	(0.056)	(0.936)	(0.470)	(0.240)
Income (logged)	-0.0539	-0.0320	0.0604***	0.100	0.265**	0.189**	0.115*
	(0.445)	(0.456)	(0.000)	(0.112)	(0.001)	(0.002)	(0.039)
Constant	1.414**	5.361***	0.686***	2.493***	0.156	-0.508	1.029*
	(0.009)	(0.000)	(0.000)	(0.000)	(0.806)	(0.277)	(0.018)
Observations	3421	3421	3421	3421	3421	3421	3421

Table 4.11 IV estimates for dilution of parent treatment.

	(1) Parent Important	(2) Parent Talks	(3) Parent Consults	(4) Parent Appreciates	(5) Parent Solves	(6) Parent trusts	(7) Parent asks opinion	(8) Parent gives reason	(9) Parent shows love
Family Size	0.0112	-0.0438	-0.0554	-0.00757	0.0404	-0.0241	**-0.0996***	**-0.0867***	-0.000504
	(0.552)	(0.184)	(0.162)	(0.810)	(0.281)	(0.458)	(0.015)	(0.037)	(0.988)
Female	0.0430**	0.0846**	0.116***	0.0569*	-0.0434	-0.0434	0.0717*	-0.0301	0.0333
	(0.009)	(0.002)	(0.001)	(0.026)	(0.156)	(0.118)	(0.030)	(0.362)	(0.226)
Age (2024)	-0.00729***	-0.00530^{+}	0.00249	-0.00898**	-0.0115***	-0.00469	-0.00808*	-0.0150***	-0.0146***
	(0.000)	(0.087)	(0.495)	(0.002)	(0.001)	(0.115)	(0.034)	(0.000)	(0.000)
First Generation	0.0597	0.0332	-0.244*	-0.0873	-0.0560	0.0167	-0.150^{+}	-0.0498	0.128^{+}
	(0.146)	(0.640)	(0.011)	(0.224)	(0.486)	(0.827)	(0.096)	(0.592)	(0.079)
Second	0.0165	-0.0361	-0.0135	-0.0326	-0.0154	-0.0657	-0.120*	-0.0316	0.0646
	(0.462)	(0.395)	(0.791)	(0.420)	(0.752)	(0.129)	(0.028)	(0.542)	(0.133)
Mother's Edu	-0.0111**	0.0119^{+}	0.0208*	-0.00389	0.0138^{+}	0.00187	0.0122	0.0235**	-0.0132^{+}
	(0.010)	(0.095)	(0.016)	(0.564)	(0.087)	(0.790)	(0.161)	(0.005)	(0.070)
Father's Edu	-0.00144	0.0104	0.0153^{+}	0.00825	0.00773	-0.00685	0.00135	0.0119	0.00328
	(0.734)	(0.158)	(0.091)	(0.236)	(0.345)	(0.342)	(0.881)	(0.164)	(0.664)
Parental ISEI	-0.000645	0.000925	0.00419**	0.00318*	0.00317*	0.00137	0.00211	0.00415*	0.00155
	(0.389)	(0.481)	(0.008)	(0.011)	(0.031)	(0.313)	(0.188)	(0.011)	(0.247)
Income (logged)	0.0549*	0.127**	0.0205	0.0658	-0.00372	0.122**	0.0882	0.0909^{+}	0.115**
	(0.028)	(0.002)	(0.683)	(0.125)	(0.937)	(0.003)	(0.102)	(0.081)	(0.006)
Constant	4.615***	2.503***	2.873***	3.524***	3.496***	3.408***	2.978***	2.915***	3.749***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3421	3421	3421	3421	3421	3421	3421	3421	3421

p-values in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.12 IV estimates for dilution of parental involvement.

	(1) Parental Interest	(2) Parental Help	(3) Parent's Evening	(4) Parent teacher conference	(5) Meets Teacher	(6) Parent representative
Family Size	**-0.0601**$^{+}$	**-0.0414***	-0.0304	**-0.0436***	-0.00317	-0.0231
	(0.067)	(0.018)	(0.109)	(0.043)	(0.863)	(0.146)
Female	-0.0808***	0.00718	-0.0301*	-0.0268	-0.0493***	0.0175
	(0.001)	(0.592)	(0.032)	(0.113)	(0.001)	(0.191)
Age (2024)	0.00107	-0.00350*	-0.00205	-0.00270	0.00218	-0.00189
	(0.709)	(0.021)	(0.205)	(0.158)	(0.170)	(0.217)
First Generation	-0.147*	-0.0436	-0.147***	0.0336	-0.0860**	-0.115***
	(0.049)	(0.266)	(0.001)	(0.460)	(0.009)	(0.000)
Second Generation	-0.000727	-0.0786***	-0.0865***	-0.00861	-0.0349	-0.0742***
	(0.985)	(0.000)	(0.000)	(0.749)	(0.111)	(0.000)
Mother's Edu (yrs.)	-0.0109	0.00289	0.0143***	-0.00461	-0.00273	0.0120**
	(0.101)	(0.410)	(0.000)	(0.324)	(0.479)	(0.003)
Father's Edu (yrs.)	-0.0186**	0.00217	0.000795	-0.00642	-0.000479	0.00749^{+}
	(0.006)	(0.546)	(0.840)	(0.174)	(0.901)	(0.075)
Parental ISEI	0.00152	0.000215	0.000744	0.000129	0.000685	0.00110^{+}
	(0.196)	(0.738)	(0.265)	(0.878)	(0.313)	(0.099)
Income (logged)	0.0772*	0.00836	-0.0146	0.0773**	0.0359	0.0400^{+}
	(0.040)	(0.719)	(0.505)	(0.002)	(0.103)	(0.055)
Constant	2.859***	0.893***	0.851***	0.303	-0.0946	-0.300^{+}
	(0.000)	(0.000)	(0.000)	(0.126)	(0.568)	(0.069)
Observations	3421	3421	3421	3421	3421	3421

p-values in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.13 Summary statistics by high, medium, and low score for cognitive ability.

Independent Variables	**High**	**Medium**	**Low**	**Total**
Family Size	**2.6656** .0384678	**2.803543** .050452	**2.892285** .043506	**2.790481** .0256974
Age Spacing	**3.496** .1009219	**3.594203** .101079	**3.770015** .100833	**3.624935** .0583682
Oldest Child	**.4368** .0198555	**.3977456** .0196561	**.4133916** .0188015	**.2058976** .0091994
Middle Child	**.1568** .0145561	**.2093398** .016339	**.2474527** .016476	**.4159338** .0112135
Youngest Child	**.4064** .0196622	**.3929147** .0196145	**.3391557** .0180754	**.3781686** .0110325
Female	**.4384** .0198635	**.4959742** .0200798	**.5269287** .0190624	**.4883601** .0113723
Age in 2024	**31.0672** .1767744	**30.91787** .1794231	**30.45269** .1715912	**30.80083** .1016581
No Migration Background	**.8768** .0131572	**.7793881** .0166531	**.6695779** .0179586	**.7718572** .009547
First Generation	**.0208** .0057131	**.0273752** .0065532	**.0480349** .0081645	**.0325918** .0040398
Second Generation	**.1024** .0121367	**.1932367** .015857	**.2823872** **.0171872**	**.195551** .0090235
Father's Education (years)	**13.8176** .1170985	**12.67874** .1118778	**11.54367** .0914478	**12.64356** .0650472
Mother's Education (years)	**13.3144** .1072198	**12.43237** .1018205	**11.20378** .0848183	**12.28091** .0597766
Highest Parental ISEI	**57.1536** .6155949	**51.11111** .6206273	**44.55313** .6102465	**50.73409** .374471
(I) Higher Managerial and Professional Workers	**.3632** .0192523	**.2302738** .0169081	**.1368268** .0131212	**.2400414** .0097171

(II) Lower Managerial and Professional Workers	**.3056** .0184412	**.3140097** .0186395	**.1892285** .0149548	**.2669426** .0100641
(IIIa) Routine Clerical Workers	**.12** .0130089	**.1400966** .0139394	**.1295488** .0128211	**.12985** .0076474
(IIIb) Routine Service and Sales Work	**.1232** .0131572	**.1304348** .0135255	**.1775837** .014591	**.1448526** .0080072
(IVa) Small Self-Employed with Employees	**.0048** .0027668	**.0161031** .0050551	**.0218341** .0055797	**.0144853** .0027183
(IVb) Small Self-Employed without Employees	**.016** .005023	**.0128824** .0045288	**.014556** .0045727	**.0144853** .0027183
(VI) Skilled Manual Workers	**.0464** .0084207	**.0901771** .0115035	**.1834061** .0147757	**.1091568** .0070945
(VIIa) Semi- and Unskilled Manual Workers	**.0192** .0054935	**.0611916** .0096258	**.1339156** .0130027	**.0734609** .0059355
(VIIb) Agricultural Labour	**--** --	**.0048309** .0027846	**.0087336** .0035525	**.004656** .0015488
(IVc) Self-Employed Farmers	**.0016** .0016	**--** --	**.0043668** .0025175	**.0020693** .0010339
Income Logged	**8.274171** .0192731	**8.134141** .0190804	**8.007135** .0171147	**8.134278** .0109333
Number of Siblings (Mother)	**2.1936** .0713528	**2.470209** .0846468	**2.935953** .0830322	**2.546301** .0467987
Number of Siblings (Father)	**2.1392** .0714482	**2.470209** .0823764	**2.778748** .0823804	**2.47284** .0460974
Age at first birth (Mother)	**26.4432** .1782409	**25.39291** .1770135	**24.09316** .1676623	**25.27056** .1028746
Age at first birth (Father)	**29.0896** .2121742	**28.2963** .2139582	**27.06405** .1908463	**28.11485** .1199363

Observations (n)	**625**	**621**	**687**	**1933**

Mean is in bold, standard errors reported beneath. Sample excludes only children.

Table 4.14 OLS estimates for sibling configuration on cognitive ability.

	(1) OLS1	(2) OLS2	(3) OLS3
Family Size (generated)	-0.394^{+}	-0.0839	-0.0104
	(0.059)	(0.649)	(0.954)
Age Spacing	-0.220*	-0.0861	-0.0728
	(0.020)	(0.326)	(0.406)
Oldest	1.104*	0.810	0.750
	(0.047)	(0.125)	(0.150)
Youngest	1.479*	1.149*	1.120*
	(0.014)	(0.046)	(0.050)
Female	-1.146**	-1.191**	-1.172**
	(0.004)	(0.002)	(0.002)
Age (2024)	0.120*	0.135**	0.161***
	(0.015)	(0.003)	(0.000)
First Generation	-4.972***	-3.385*	-2.960*
	(0.001)	(0.011)	(0.026)
Second Generation	-4.746***	-2.386***	-2.303***
	(0.000)	(0.000)	(0.000)
Number of Years of Education (Father)		0.605***	0.347***
		(0.000)	(0.001)
Number of Years of Education (Mother)		0.757***	0.623***
		(0.000)	(0.000)
Parental ISEI			0.0742***
			(0.000)
Income (logged)			0.758
			(0.139)
Constant	30.56***	11.55***	5.420
	(0.000)	(0.000)	(0.225)
Observations	1933	1933	1933

p-values in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.15 First stage IV results for cognitive ability.

	IV1: First-Stage Regression	IV2: First-Stage Regression	IV3: First-Stage Regression
No. Of Siblings (Mother)	0.0415*	0.0471*	0.0452*
	(0.040)	(0.024)	(0.036)
No. Of Siblings (Father)	0.0422**	0.0461**	0.0438**
	(0.005)	(0.003)	(0.006)
Mother's age at first birth	-0.0610***	-0.0666***	-0.0663***
	(0.000)	(0.000)	(0.000)
Father's age at first birth	0.00444	0.00394	0.00306
	(0.537)	(0.579)	(0.660)
Age Spacing	-0.0501***	-0.0498***	-0.0483***
	(0.000)	(0.000)	(0.000)
Oldest	-1.088***	-1.082***	-1.067***
	(0.000)	(0.000)	(0.000)
Youngest	-1.159***	-1.158***	-1.151***
	(0.000)	(0.000)	(0.000)
Female	-0.0471	-0.0490	-0.0525
	(0.268)	(0.247)	(0.219)
Age (2024)	-0.0230***	-0.0240***	-0.0195**
	(0.000)	(0.000)	(0.005)
First Generation	0.359	0.371	0.347
	(0.222)	(0.204)	(0.213)
Second Generation	0.112	0.137	0.130
	(0.264)	(0.167)	(0.179)
Mother's Edu (yrs.)		0.0201	0.0188
		(0.116)	(0.151)
Father's Edu (yrs.)		0.00669	0.00810
		(0.492)	(0.449)
Parental ISEI			-0.00450
			(0.123)
Logged Income			0.194*
			(0.036)
Constant	5.846***	5.667***	4.192***
	(0.000)	(0.000)	(0.000)
Observations	1933	1933	1933

Table 4.16 Second stage IV estimates for sibling configuration on cognitive ability.

	(1) IV1	(2) IV2	(3) IV3
Family Size (generated)	-6.797***	-1.679*	-0.907
	(0.000)	(0.033)	(0.247)
Age Spacing	-0.411***	-0.140	-0.102
	(0.000)	(0.137)	(0.270)
Oldest	-7.119***	-1.205	-0.359
	(0.000)	(0.284)	(0.741)
Youngest	-6.996***	-0.926	-0.0321
	(0.000)	(0.430)	(0.978)
Female	-1.446**	-1.263**	-1.217**
	(0.004)	(0.001)	(0.001)
Age (2024)	0.0438	0.115*	0.153***
	(0.473)	(0.012)	(0.001)
First Generation	-1.433	-2.591+	-2.563+
	(0.579)	(0.089)	(0.072)
Second Generation	-2.649**	-1.986***	-2.093***
	(0.002)	(0.001)	(0.000)
Number of Years of Education (Father)		0.721***	0.610***
		(0.000)	(0.000)
Number of Years of Education (Mother)		0.573***	0.345***
		(0.000)	(0.001)
Parental ISEI			0.0661***
			(0.000)
Income (logged)			0.886
			(0.103)
Constant	58.21***	19.31***	8.783+
	(0.000)	(0.000)	(0.088)
Observations	1933	1933	1933

p-values in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.17 Hansen-J statistic for each dependent variable.
Chi-squared p-values reported in parentheses. Variables in bold represent cases in which at least one instrument is correlated with the error term. Variables in italics represent cases in which the p-value is only marginally larger than 0.05.

Dependent Variable	Hansen-J
Gymnasium IV1	*6.461* (*0.0912*)
Gymnasium IV2	**7.446 (0.0590)**
Gymnasium IV3	**8.921 (0.0304)**
Reading	**10.363 (0.0157)**
TV	*6.458 (0.0913)*
Own Room	*6.583 (0.0864)*
Sport	**10.906 (0.0122)**
Music	**12.018 (0.0073)**
Dance	2.518 (0.4720)
Tech-Work	2.025 (0.5673)
Parent important	1.086 (0.7804)
Parent talks	3.348 (0.3410)

Parent consults	*6.630* *(0.0847)*
Parents express appreciation	**8.033** **(0.0453)**
Parents able to solve problems with you	0.583 (0.9004)
Parents have impression of trusting you	4.790 (0.1878)
Parents asks opinion on family matter	2.976 (0.3953)
Parents give reason for making decision	3.474 (0.3241)
Parents show they love you	3.103 (0.3761)
Parents show interest in performance	3.685 (0.2976)
Parents help with studying	1.564 (0.6677)
Parents take part in parent's evening	0.492 (0.9208)
Parents come to teacher's office hours	**11.750** **(0.0083)**
Parents visit teacher outside office hours	1.826 (0.6094)
Parents involved as a representative	1.692 (0.6387)
Cognitive Ability IV1	0.435 (0.9329)

Cognitive Ability IV2	0.693 (0.8750)
Cognitive Ability IV3	0.689 (0.8759)

Table 4.18 Ordered logit and logit results for environment and opportunities.

	(1) Reading	(2) TV	(3) Own room	(4) Sport	(5) Music	(6) Dance	(7) Tech work
Family Size	**0.0588***	**-0.267*****	**-0.313*****	0.0181	**0.0801****	-0.0276	**-0.0911****
	(0.024)	(0.000)	(0.000)	(0.518)	(0.006)	(0.356)	(0.004)
Female	0.569***	-0.294***	0.0233	-0.477***	0.204**	0.991***	-1.446***
	(0.000)	(0.000)	(0.868)	(0.000)	(0.002)	(0.000)	(0.000)
Age (2024)	0.0252***	0.0315***	-0.0426**	-0.0272***	-0.0346***	0.0220***	0.0370***
	(0.000)	(0.000)	(0.009)	(0.000)	(0.000)	(0.001)	(0.000)
First Generation	0.545***	-0.313	-1.101***	0.241	-0.109	-0.234	-0.259
	(0.000)	(0.232)	(0.000)	(0.166)	(0.526)	(0.139)	(0.162)
Second Generation	0.0942	-0.338**	-1.500***	-0.0590	-0.0279	0.0167	-0.0407
	(0.324)	(0.005)	(0.000)	(0.571)	(0.790)	(0.869)	(0.695)
Father's Edu (yrs.)	0.0960***	-0.0332	-0.0289	0.0426*	0.0836***	0.0176	-0.0166
	(0.000)	(0.138)	(0.573)	(0.016)	(0.000)	(0.311)	(0.334)
Mother's Edu (yrs.)	0.0458**	-0.0377^{+}	0.0892^{+}	0.0559**	0.0305^{+}	0.0570***	0.0109
	(0.006)	(0.079)	(0.075)	(0.001)	(0.078)	(0.001)	(0.513)
Parental ISEI	0.00564*	-0.00740*	0.0194**	0.00250	0.00238	0.00514^{+}	0.00548^{+}
	(0.044)	(0.049)	(0.005)	(0.412)	(0.414)	(0.080)	(0.073)
Logged Income	-0.0890	-0.152	0.701***	0.219*	0.239*	0.225*	0.104
	(0.313)	(0.212)	(0.000)	(0.017)	(0.011)	(0.014)	(0.228)
Constant			-2.107				
			(0.221)				
/ cut1	0.449	-7.452***		-0.197	2.664***	4.116***	1.376^{+}
	(0.540)	(0.000)		(0.794)	(0.001)	(0.000)	(0.057)
cut2	2.094**	-5.609***		0.759	3.327***	5.191***	2.491***
	(0.004)	(0.000)		(0.316)	(0.000)	(0.000)	(0.001)
cut3	2.650***	-5.297***		1.036	3.486***	5.694***	3.065***
	(0.000)	(0.000)		(0.171)	(0.000)	(0.000)	(0.000)
cut4	3.795***	-3.604***		3.484***	4.430***	7.869***	4.654***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
Observations	3421	3421	3421	3421	3421	3421	3421

Table 4.19 Ordered logit and logit and results for parental treatment.

	(1) Parent Important	(2) Parent Talks	(3) Parent consults	(4) Parent appreciates	(5) Parent solves	(6) Parent trusts	(7) Parent asks opinion	(8) Parent reason	(9) Parent shows love
Family Size	-0.0561^{+}	**-0.0789****	**-0.0927*****	**-0.0782****	-0.0177	**-0.0580***	**-0.135*****	-0.0465	**-0.0845***
	(0.079)	(0.004)	(0.000)	(0.006)	(0.507)	(0.042)	(0.000)	(0.141)	(0.012)
Female	0.195**	0.175**	0.207***	0.135*	-0.0877	-0.0705	0.117^{+}	-0.0917	0.121^{+}
	(0.008)	(0.004)	(0.001)	(0.025)	(0.149)	(0.254)	(0.053)	(0.143)	(0.055)
Age (2024)	-0.0385***	-0.0117^{+}	0.00947	-0.0219***	-0.0233***	-0.00861	-0.0109	-0.0201**	-0.0336***
	(0.000)	(0.076)	(0.139)	(0.001)	(0.000)	(0.183)	(0.105)	(0.003)	(0.000)
First Generation	0.448*	0.0297	-0.461**	-0.169	-0.0530	0.0528	-0.289^{+}	-0.145	0.352*
	(0.020)	(0.841)	(0.005)	(0.303)	(0.716)	(0.741)	(0.060)	(0.353)	(0.027)
Second Generation	0.0570	-0.0480	-0.0554	-0.0753	0.0315	-0.123	-0.226*	-0.0972	0.228*
	(0.571)	(0.589)	(0.518)	(0.408)	(0.732)	(0.192)	(0.020)	(0.280)	(0.019)
Father's Edu (yrs.)	-0.00744	0.0228	0.0239	0.0181	0.0105	-0.0157	0.0000494	0.0231	0.0111
	(0.701)	(0.164)	(0.143)	(0.280)	(0.528)	(0.342)	(0.998)	(0.160)	(0.529)
Mother's Edu (yrs.)	-0.0438*	0.0245	0.0368*	-0.00737	0.0276^{+}	0.0102	0.0185	0.0425**	-0.0291^{+}
	(0.016)	(0.118)	(0.021)	(0.650)	(0.092)	(0.535)	(0.248)	(0.009)	(0.087)
Parental ISEI	-0.00338	0.00286	0.00745**	0.00597*	0.00554^{+}	0.00275	0.00367	0.00888**	0.00264
	(0.306)	(0.286)	(0.006)	(0.035)	(0.052)	(0.347)	(0.192)	(0.002)	(0.362)
Logged Income	0.290**	0.257**	0.0604	0.204*	0.0604	0.280**	0.180^{+}	0.166^{+}	0.305**
	(0.009)	(0.002)	(0.487)	(0.047)	(0.499)	(0.001)	(0.064)	(0.080)	(0.001)
/									
cut1	-7.194***	-2.961***	-2.269**	-5.148***	-3.862***	-3.996***	-2.459**	-2.138**	-4.767***
	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.002)	(0.006)	(0.000)
cut2	-6.501***	-2.036**	-1.740*	-4.636***	-3.367***	-3.463***	-2.052**	-1.715*	-4.323***
	(0.000)	(0.004)	(0.017)	(0.000)	(0.000)	(0.000)	(0.009)	(0.028)	(0.000)
cut3	-4.090***	-0.434	-0.420	-2.496**	-1.878*	-1.605*	-0.643	-0.301	-2.625***
	(0.000)	(0.530)	(0.561)	(0.003)	(0.011)	(0.025)	(0.414)	(0.698)	(0.001)
cut4	-3.211***	0.426	0.0524	-1.801*	-1.387^{+}	-1.053	-0.246	0.166	-1.815*
	(0.000)	(0.536)	(0.942)	(0.030)	(0.062)	(0.139)	(0.754)	(0.830)	(0.019)

Table 4.20 Order Logit and Logit results for parental involvement.

	(1) Parental Interest	(2) Parent Helps	(3) Parent's Evening	(4) Parent Conference	(5) Meets Teacher	(6) Parent Representative
main						
Family Size	**-0.128*****	**-0.106****	**-0.140*****	-0.0338	0.00497	0.0270
	(0.000)	(0.001)	(0.000)	(0.289)	(0.894)	(0.572)
Female	-0.213**	0.0420	-0.179*	-0.115^{+}	-0.293***	0.116
	(0.001)	(0.620)	(0.032)	(0.097)	(0.001)	(0.206)
Age (2024)	0.00436	-0.0174^{+}	-0.0105	-0.00673	0.0135	-0.00820
	(0.546)	(0.051)	(0.247)	(0.364)	(0.125)	(0.407)
First Generation	-0.385^{+}	-0.330^{+}	-0.704***	0.0335	-0.622*	-2.046***
	(0.052)	(0.088)	(0.000)	(0.845)	(0.015)	(0.000)
Second Generation	-0.00298	-0.485***	-0.431***	-0.0928	-0.226^{+}	-0.820***
	(0.976)	(0.000)	(0.000)	(0.376)	(0.091)	(0.000)
Father's Edu (yrs.)	-0.0519**	0.0135	0.00487	-0.0274	-0.00335	0.0382
	(0.003)	(0.573)	(0.845)	(0.159)	(0.880)	(0.134)
Mother's Edu (yrs.)	-0.0269	0.0217	0.0926***	-0.0159	-0.0156	0.0737**
	(0.125)	(0.350)	(0.000)	(0.405)	(0.483)	(0.001)
Parental ISEI	0.00410	0.00284	0.00416	0.00226	0.00431	0.0112*
	(0.173)	(0.469)	(0.275)	(0.495)	(0.265)	(0.011)
Logged Income	0.207*	-0.00734	-0.0898	0.254*	0.202	0.174
	(0.027)	(0.959)	(0.488)	(0.012)	(0.105)	(0.201)
Constant		1.795	1.497	-0.927	-3.150**	-4.627***
		(0.114)	(0.156)	(0.260)	(0.002)	(0.000)
/						
cut1	-3.642***					
	(0.000)					
cut2	-0.692					
	(0.380)					
cut3	1.665*					
	(0.035)					
Observations	3421	3421	3421	3421	3421	3421

Figure 4.1 OLS results for probability of attending Gymnasium.

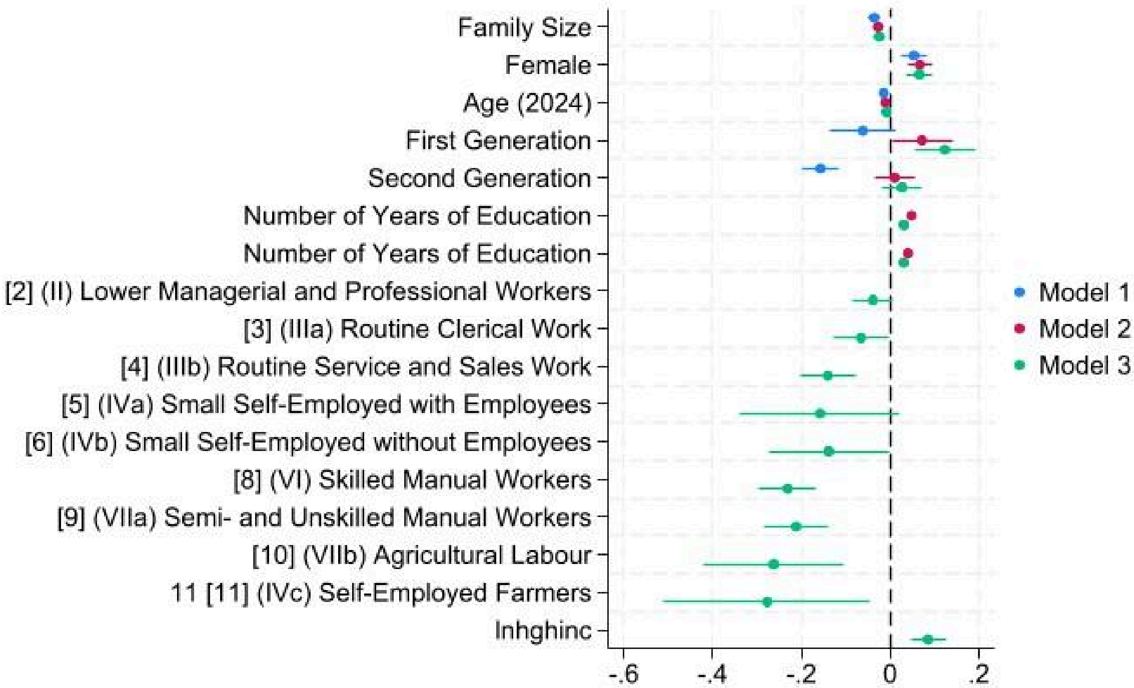


Figure 4.2 IV stage two results for probability of attending Gymnasium.

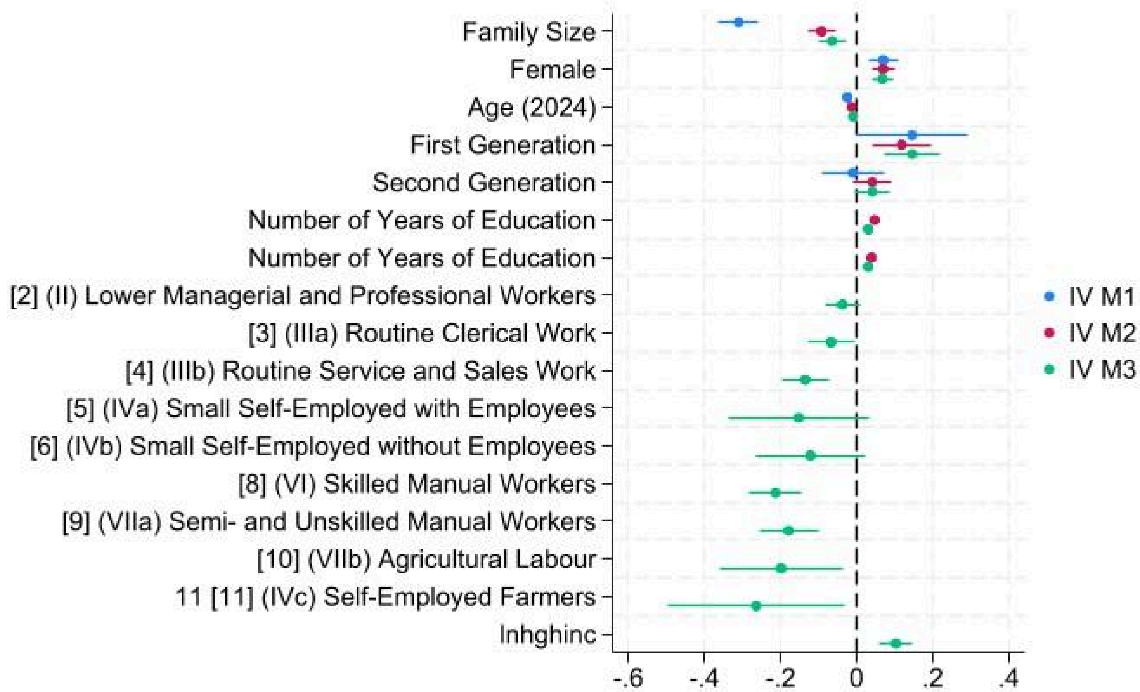
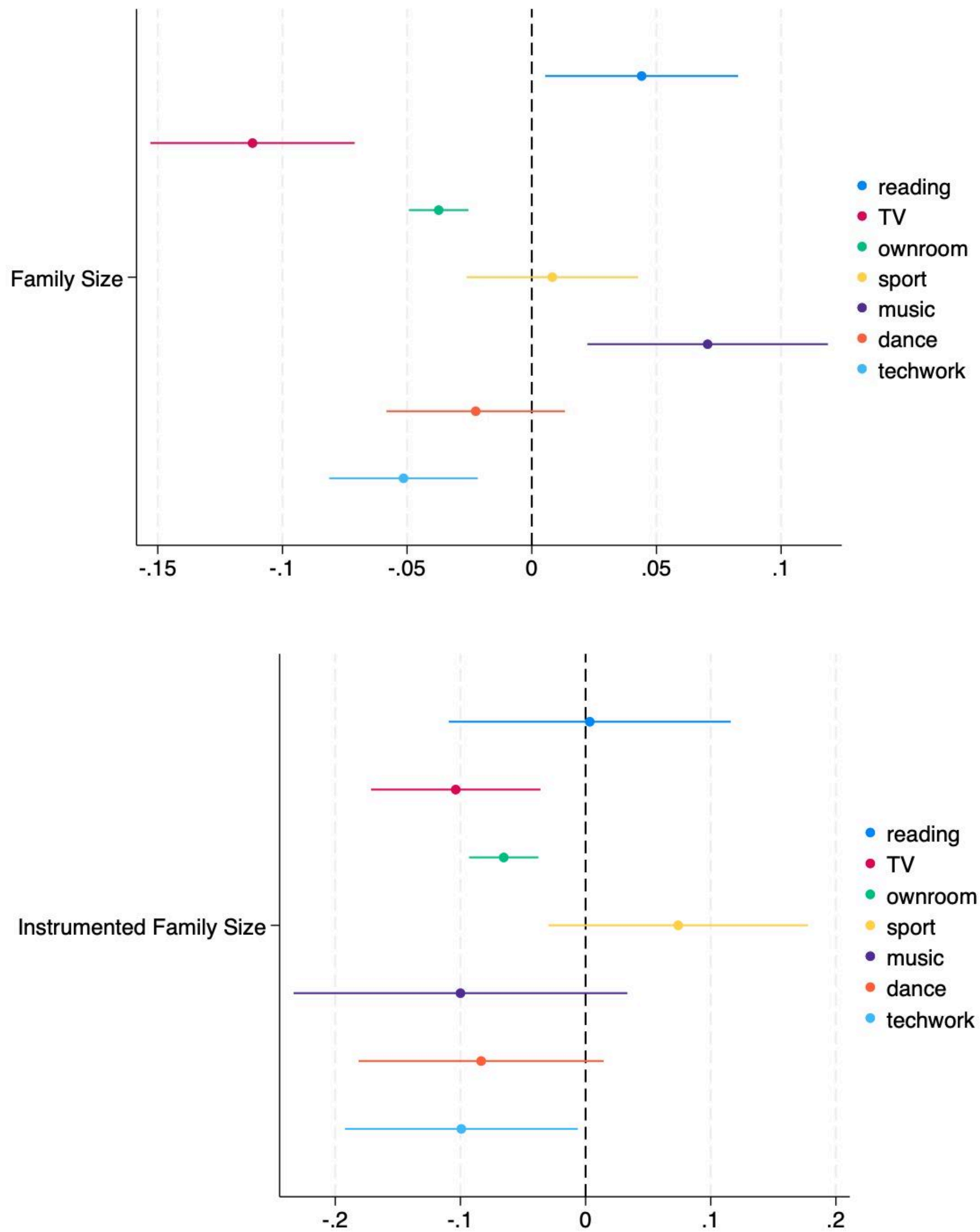
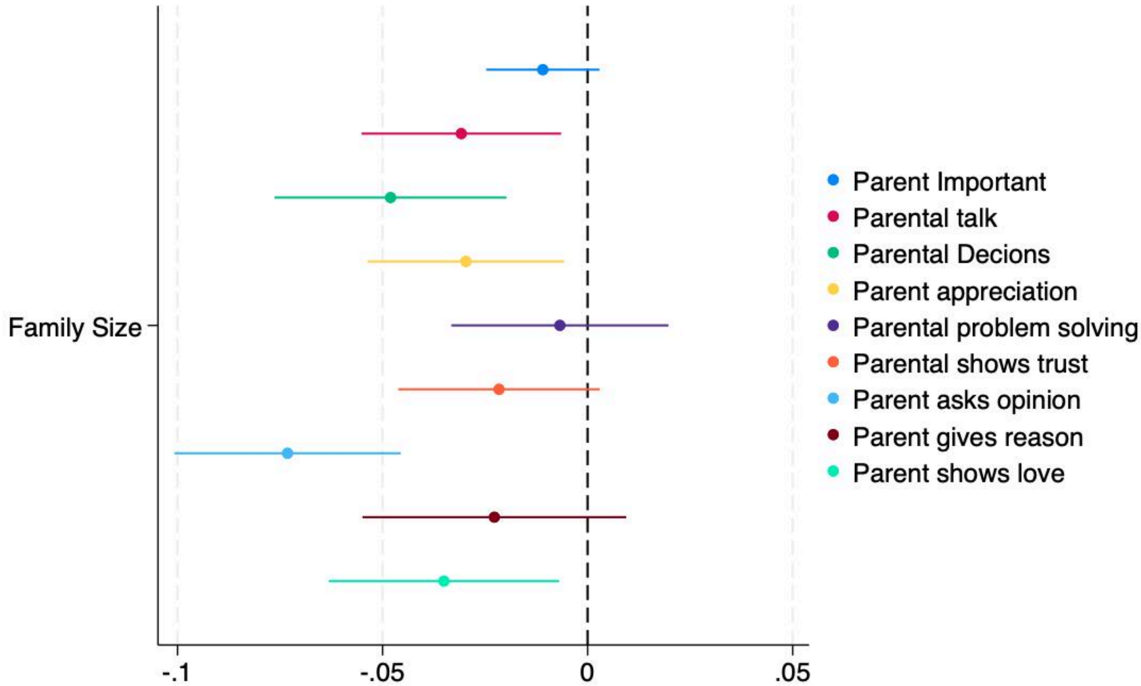


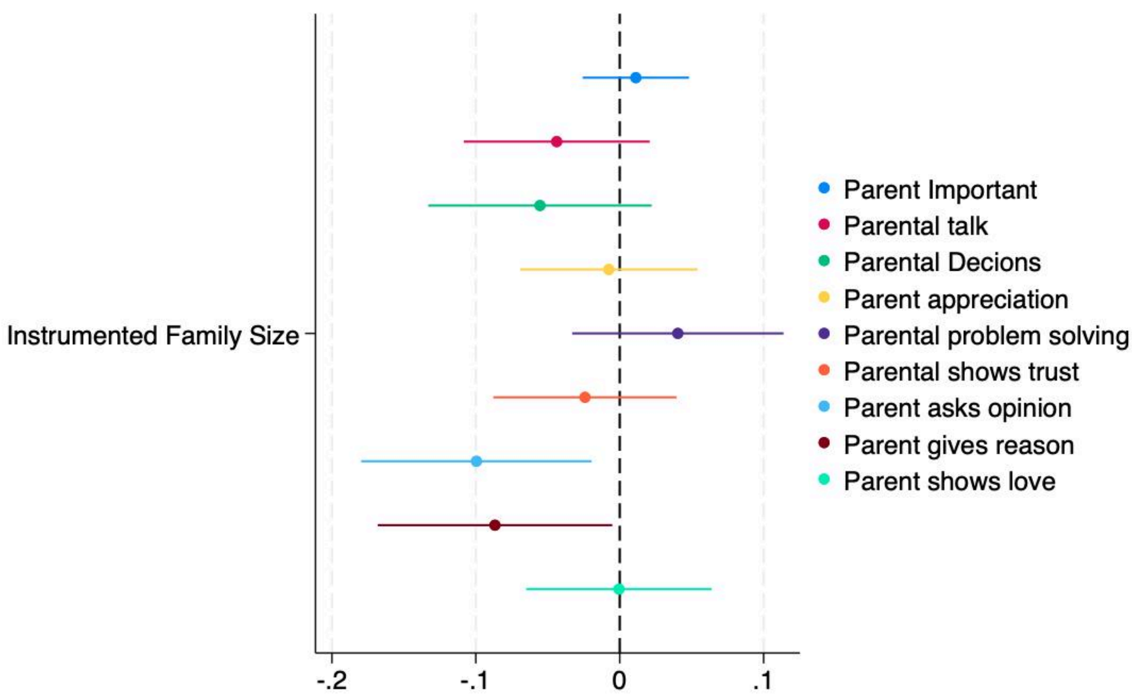
Figure 4.3 A comparison of OLS and IV estimates on environment and opportunities.



Controls included in each model but not depicted.

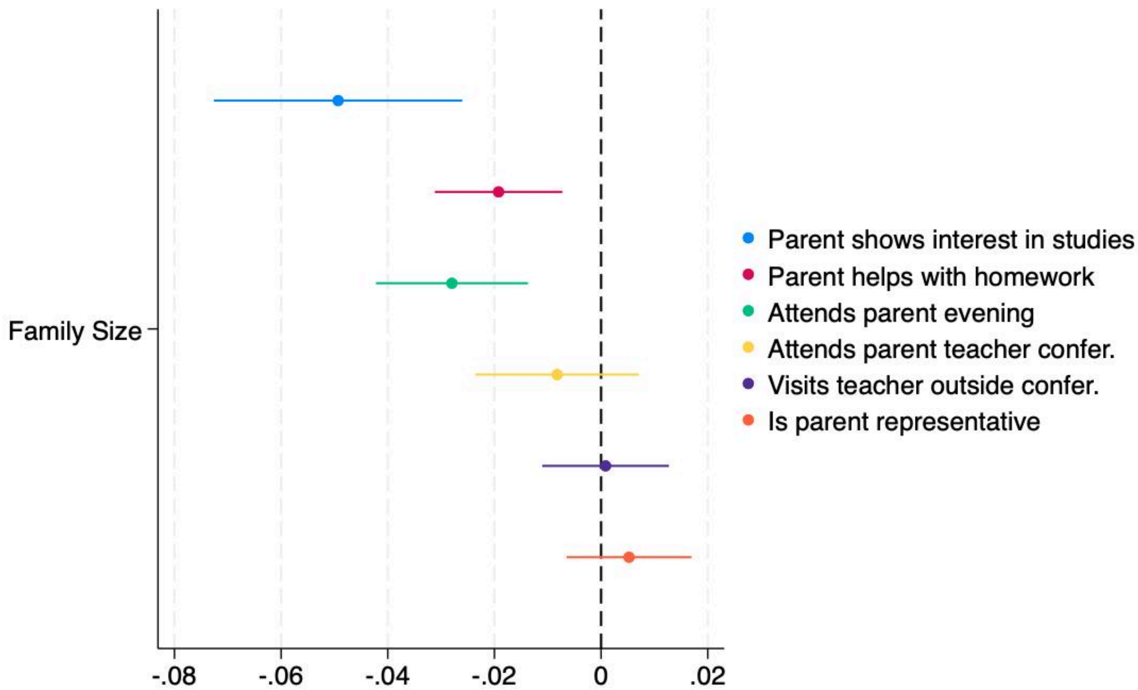
Figure 4.4 A comparison of OLS and IV estimates on parental treatment by family size.

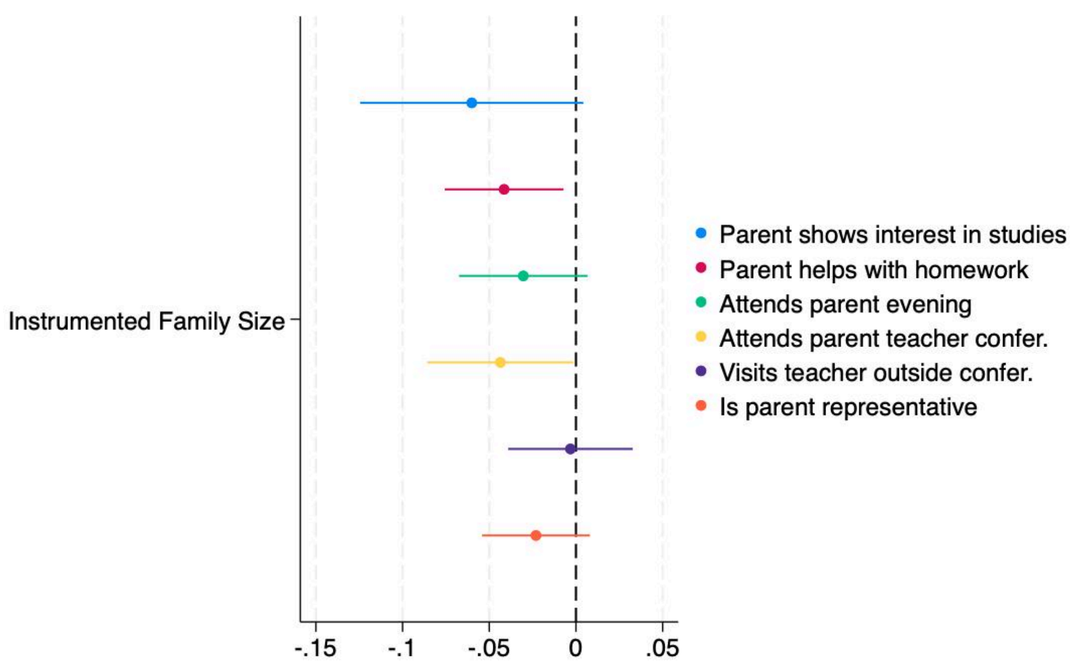




Controls included in each model but not depicted.

Figure 4.5 A comparison of OLS and IV estimates on parental involvement by family size.





Controls included in each model but not depicted.