

# **Family Behavior and Children's Wellbeing: Statistical Modeling and Measurement Issues**

Dissertation

zur Erlangung des akademischen Grades  
**Doctor of Philosophy (Ph.D)**

im Fach  
**Soziologie**

Tag der Verteidigung 16.12.2022

an der Kultur-, Sozial- und Bildungswissenschaftlichen Fakultät der  
Humboldt-Universität zu Berlin

von **Alejandra Rodríguez Sánchez**

Prof. Dr. Julia von Blumenthal  
Präsidentin der  
Humboldt-Universität zu Berlin

Prof. Dr. Christian Kassung  
Dekan der Kultur-, Sozial- und  
Bildungswissenschaftlichen Fakultät

Gutachterin /Gutachter:

1. Prof. Dr. Anette Eva Fasang
2. Prof. Dr. Juho Härkönen
3. Prof. Dr. Philipp Lersch

## Acknowledgments

I could not have undertaken this journey without the encouragement, support, and guidance of Prof. Dr. Anette Eva Fasang, sincerely one of the best mentors one could possibly wish for. I can only hope our intellectual exchange will extend further into the future.

I am extremely grateful to the Rosa-Luxemburg-Stiftung. Without their funding for the first years of my doctoral studies, this work could not have been started.

Special thanks to Jascha Dräger, for being one of the toughest critics one could wish for at every stage of research, but especially for his encouragement in the times I dealt with the often experienced rejection of my papers at the first, the second, and even the third attempt at publication.

Thanks to the many participants in the Social Demography Writing Workshop, I hope their critical spirit is reflected in all the chapters of this dissertation. I learned a lot about writing from the collective intelligence that has accumulated there over the years.

Thanks to my colleagues at the BGSS – Ana Luisa Azenha, Claire Samtleben, Gabriela Camacho, Mahmoud Farag, Pablo Beytía, and Philippe Joly – for the company, laughter, and advice during the years we spent writing together.

Lastly, I would like to mention my mother, Luz Dary Sánchez Ramírez, and sister, Daniela Suárez Sánchez, who, despite not knowing much about family sociology, were avid listeners to all I had to tell them about my work. Our unconventional family life - including our feline relatives - is what keeps me going.

# Contents

**Acknowledgments**

**List of Figures**

**List of Tables**

**INTRODUCTION** **i**

Theoretical Background . . . . .	iv
Summary of chapters . . . . .	xi

**CHAPTER 2 - Fair comparisons: life course selection bias and  
the effect of father absence on US children** **1**

Background . . . . .	4
Father absence in US family sociology: Which causal effects?	4
Life course selection bias . . . . .	8
Methods and Data . . . . .	10
Causal inference for family transition effects, re-examined . .	10
Bayesian additive regression trees . . . . .	13
Data . . . . .	16
Variables: treatment, confounders, and outcomes . . . . .	16
Analytical strategy . . . . .	23
Results . . . . .	25
Discussion . . . . .	29
Tables and Figures . . . . .	38

**CHAPTER 3 - The complex effects of family instability on  
adolescent problem behavior in a U.S. birth cohort** **50**

Background . . . . .	54
Problem behavior in a cohort of American children . . . . .	54
Family instability: the accumulation and timing of its effects	55

Dynamic selection into family instability from a life-course perspective . . . . .	57
Methods and Data . . . . .	59
Methods . . . . .	59
Data . . . . .	65
Results . . . . .	69
Discussion . . . . .	72
Tables and Figures . . . . .	77

**CHAPTER 4 - What can parents do? The causal mediating role of parenting in explaining SES differences in children’s language development 98**

Background . . . . .	100
Parenting and the development of language skills . . . . .	102
Differences in parenting behavior by SES . . . . .	103
SES as a causal factor in parenting and children’s language development . . . . .	106
Data and Methods . . . . .	109
Causal interventional direct and indirect effects . . . . .	109
<i>Data</i> . . . . .	116
Exposure: Latent class analysis to measure SES . . . . .	117
<i>Outcome: language skills</i> . . . . .	119
<i>Mediators: Parenting</i> . . . . .	120
Exposure-induced confounders of the mediator-outcome association . . . . .	122
Results . . . . .	124
Discussion . . . . .	127
Tables and Figures . . . . .	133

**CHAPTER 5 - Is it just noise? Measuring unobservable cogni-**

<b>tive abilities in early childhood</b>	<b>147</b>
Background . . . . .	150
Pitfalls of the traditional validation framework . . . . .	150
Ordinal v. Interval: properties of scales . . . . .	152
Measurement invariance as measurement error . . . . .	153
Methods . . . . .	157
Data . . . . .	157
Methods . . . . .	159
Results . . . . .	164
Discussion . . . . .	168
Tables and Figures . . . . .	174
<b>CONCLUSION</b>	<b>179</b>
Policy implications . . . . .	185
Limitations . . . . .	186
Discussion . . . . .	188
<b>APPENDIX</b>	<b>192</b>
Supplementary Materials - Chapter 2 . . . . .	192
Part 1 - Bayesian Additive Regression Trees . . . . .	192
Part 2 - Additional Tables with descriptive statistics . . . . .	194
Supplementary Materials - Chapter 4 . . . . .	230
Descriptive statistics . . . . .	230
Clarifications on causal mediation analysis . . . . .	233
Robustness check employing mother's educational attainment	
instead of latent class approach to SES . . . . .	234
Supplementary Materials - Chapter 5 . . . . .	235
Sample Descriptive Statistics . . . . .	235
Latent Class Analysis for the Social Class Structure . . . . .	235

Parametric Item Response Theory (PIRT): partial credit model for polytomous item responses . . . . .	236
Tables and Figures . . . . .	239
<b>Bibliography</b>	<b>247</b>

## List of Figures

1	Figure 1: Sample Sizes at Each Wave by Family Instability, FFCWS . . . . .	43
2	Figure 2: CATE of Departure of Bio. Father on Health wellbeing Outcomes, FFCWS . . . . .	44
3	Figure 3: CATE of Departure of Bio. Father on Behavioral/Noncognitive wellbeing Outcomes, FFCWS . . . . .	45
4	Figure 4: CATE of Departure of Bio. Father on Cognitive/Schooling wellbeing Outcomes, FFCWS . . . . .	46
5	Figure 5: Timing - CATE of Earlier Father Departure on Later Early Childhood wellbeing Outcomes, FFCWS . . . . .	47
6	Figure 6: Timing - CATE of Earlier Father Departure on Later Middle Childhood wellbeing Outcomes, FFCWS . . . . .	48
7	Figure 7: Timing - CATE of Earlier Father Departure on Later Adolescence wellbeing Outcomes, FFCWS . . . . .	49
8	Figure 8: Simplified DAG for the family instability hypothesis and the Selection Hypotheses . . . . .	93
9	Figure 9: Average problem behavior at various ages following all potential exposures . . . . .	94
10	Figure 10: Estimated cumulative effects of repeated family changes on problem behavior at 15 years old . . . . .	95
11	Figure 11: Moderating effect of cumulative changes in family structure on problem behavior at age 15 by child's mother racial-ethnic category and child's gender assigned at birth . . . . .	96
12	Figure 12: Joint effect family instability exposure trajectories up to 15 years old on the internalizing, externalizing and total problem behavior of adolescents [ref. stable families (0,0,0,0)] . . . . .	97
13	Figure 13: DAG representation of the theoretical associations between parenting mediators and exposure-induce confounding	145

14	Figure 14: Interventional/Randomized mediation effects by SES on language skills at three time points with respect to High SES parents . . . . .	146
15	Figure 15: Directed Acyclic Graph Illustrating Measurement Error Caused by Measurement Invariance . . . . .	175
16	Figure 16: Mathematics Rest-score Relative Risk Adjusted by Rest Score . . . . .	176
17	Figure 17: Relative risks for Various Covariates Adjusted by Rest Score . . . . .	177
18	Figure 18: Standardized Estimated Coefficients in Linear Model (Panel A) and Average Marginal Effects from Cumulative Link Model (Panels B and C) for Scores on Mathematics Test . . . . .	178
S1	Figure S 1: Interventional/Randomized mediation effects by maternal education (ISCED) on language skills at three time points with respect to mothers with university education . . .	234
S2	Figure S 2: Polychoric Correlation for Mathematics Test Items	243
S3	Figure S 3: Person and Item Fit Plot from Graded Response Model in Mathematics Test Items . . . . .	245
S4	Figure S 4: Number of Gutman Errors and Number of Children in the Mathematics Test Items and their Mokken Subscale . .	246

## List of Tables

1	Description of variables used in chapter 2 . . . . .	39
2	Description of variables used in chapter 3 . . . . .	78
3	Average score on items and subscales of problem behavior at fifteen-year follow-up and descriptives by gender and racial- ethnic background . . . . .	84
4	Descriptive Statistics for Analytical Variables in the FFCWS Initial Sample by Family Structure at Birth . . . . .	85
5	Description of variables used in Chapter 4 . . . . .	134
6	Composition of Socioeconomic Status Latent Classes by Moth- ers and Fathers' Characteristics, NEPS SC1 . . . . .	142
7	Parenting mediators and exposure-induced confounders by SES, NEPS SC1 . . . . .	143
8	ACM Checks for Mathematics Test, PPVT-4 and ELFRA-2P	174
9	Mokken Scale Analysis Results for Mathematics Test Items .	175
10	Number and Proportion of Items with DIF by Generalized Mantel-Haezel Test . . . . .	175
11	Effect Size Estimates from Linear Models in Panel A . . . . .	176
S1	Descriptive Statistics for Analytical Variables in the FFCWS at Baseline by Type of Transition at each wave (baseline, 2nd and 3rd waves) . . . . .	195
S2	Descriptive Statistics for Analytical Variables in the FFCWS at Baseline by Type of Transition at each wave (4th - 6th waves)	201
S3	Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 3 . . . . .	207
S4	Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 4 . . . . .	210
S5	Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 5 . . . . .	215

S6	Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 6 . . . . .	222
S7	Descriptive Statistics by Parents' Socioeconomic Status, NEPS SC1 . . . . .	231
S8	Sample Descriptive Statistics Complete Cases in NEPS SC1 5th Wave . . . . .	240
S9	Sample Descriptive Statistics Complete Cases by Standardized Test in NEPS SC1 . . . . .	241
S10	Latent Class Analysis for Social Class Structure in NEPS SC1 Cohort at Wave 1 . . . . .	242
S11	Mathematics Test Items . . . . .	243
S12	Internal Reliability Coefficients for Mathematics Test Items .	244
S13	Criteria for Number of Factors in Mathematics Test Items . .	244
S14	Confirmatory Factor Analysis for Mathematics Test Items . .	244
S15	Full-Information Item Bi-factor and Two-Tier Analysis for Mathematics Test Items . . . . .	244
S16	Graded Response Model Estimates for Mathematics Test Items	245
S17	AISP Genetic Algorithm for Mathematics Test Items . . . . .	246
S18	Mathematics Items Flagged as DIF by Generalized Mantel- Haezel Test . . . . .	247

## INTRODUCTION

Contemporary sociological studies on the family are focused on understanding the causes and effects of specific family behaviors (Bianchi 2014). For example, among some of these behaviors, research has examined the effects of marriages, unions, fertility, parenting, divorce, and separation on adults and children (see Crosnoe et al. 2014; Smock and Schwartz 2020). Questions of interest are: What factors affect these behaviors? And what are the consequences these behaviors may have for adults' and children's wellbeing? One salient perspective in family sociology considers families as involved in a complex process of transmitting the conditions that make possible certain standards of living and demographic behaviors from one generation to the next (Edin and Kissane 2010; Amato and Patterson 2017). From this point of view, analogous to an inheritance (Bourdieu 1993), parents pass on to their children – the next generation – not just the resources that were the product of their parents' work, but also the ways of being, the *habitus*, expectations, and all sorts of cultural behaviors that may be relevant explanatory factors for the transmission of social inequality.

The study of the association between family behavior and children's wellbeing follows a long tradition in demographic and sociological research (see Bianchi 2014; Seltzer 2019; Coontz 2016). The first studies on this topic were focused on the observed differences in family behavior among individuals of different socioeconomic backgrounds. One area of research of continuing focus is potentially negative effects on children of the demographic behavior of their parents. For example, early studies of marriage patterns by social class in Wales and England in the 19th century were focused on the fertility behavior of working-class families, their family size, and what this entailed for infant child mortality (for a summary see Garrett et al. 2001). Differences in family behavior by social class were

negatively valued and considered problematic from an early social policy, anti-poverty and charity based, perspective. Later on, in the US, and continuing this line of research, demographic studies diagnosed the family structures of Blacks - where a higher prevalence of non-nuclear family structures was observed - as a likely explanatory factor of the multiple social problems among this group in comparison to Whites (see Du Bois 1899; Frazier 1928). One could say that, although these early studies did not state their claims in these terms, these were first attempts at studying how demographic family behavior could shape – through its potential unintended negative effects on children – the inter-generational reproduction of social inequality.

The long term changes in family behavior observed by scholars during the 20th and 21st centuries led many researchers to explore what potential consequences such drastic, rapid changes might have on the wellbeing of children (see F. Bernardi and Boertien 2017; Hadfield et al. 2018). Such claims led to the hypothesis that changes in family behavior were, at least partly, an explanation for the lower educational and occupational attainment of children coming from specific family backgrounds, given that certain backgrounds made children more prone to experiencing disruptions in parental family life courses (Fomby and Cherlin 2007). And, therefore, an explanatory factor of the reproduction of social inequality through the diverging destinies of children growing up in non-nuclear families (see Bloome 2017; McLanahan 2004; McLanahan and Percheski 2008).

But is that so? Can we claim that parents' demographic behavior negatively affects the chances of children? In this dissertation, I examine often studied mechanisms linking family behavior, children's wellbeing, and social inequality: the “negative” effects of divorce and family instability on children, and the degree to which parenting practices, activities, and styles can explain socioeconomic status (SES) differences in cognitive development

among children (see Wu and Martinson 1993; Cavanagh and Fomby 2019; Nomaguchi and Milkie 2020), but paying close attention to measurement and modeling issues that arise when examining unobservable variables such as children’s cognitive development. Establishing cause and effect relations in the study of the intergenerational transmission of inequality is complicated by the fact that family behavior is endogenous (Ginther and Pollak 2004), meaning, that is itself an effect of events that took place in the recent or distant past, and a cause of the family behavior that will take place in the future.

The reason why I write “negative” effects is that quantifying specific dimensions of children’s wellbeing, in particular those connected to educational and occupational attainment, such as cognitive development, ought to be reconsidered in the light of recent advances in causal inference and measurement theories in psychology. Family instability, in turn, also requires closer scrutiny. This is a broad concept capturing the notion of repeated family transitions, closely related to other demographic concepts, such as family complexity (Smock and Schwartz 2020; Van Winkle 2018). In basic terms, family instability refers to repeated family changes caused by disruptions in the romantic relationships of adults and which are experienced by children (Cavanagh and Fomby 2019). Although children can experience multiple other family changes, divorce or separation play a prominent role in this literature (McLanahan, Tach, and Schneider 2013). In the context of the second demographic transition (Lesthaeghe 1991; Schweizer 2020), parental separation is the most frequent type of transition children experience. For this reason, it is reasonable to examine this specific transition, without disregarding the broader concept of family instability and the timing of the effects of family changes.

Finally, lying in between family instability and children’s wellbeing are multiple mechanisms that presumably link parents and their parenting –

another type of family behavior occurring within families – with child development (Kalil, Ryan, and Corey 2012). These mechanisms also require a more throughout examination in terms of causal inference. Parenting, broadly understood, describes the type of activities parents do with their children with, among others, the purpose of enhancing their potential to thrive as adults (Nomaguchi and Milkie 2020). However, parenting behavior varies substantially by parental SES (Greg J. Duncan, ZiolGuest, and Kalil 2010). One of the explanations for why children exposed to family instability perform worse than children in stable families has to do with the parenting of lone mothers and the socialization of their children (Cavanagh and Fomby 2019; Grusec 2011), though little evidence support parenting as a mediating mechanism of family changes exist (for some findings see Brand et al. 2019). The reason for this general lack of evidence might lie in that parenting activities have not been studied under a causal inference perspective, in what is known as causal mediation analysis. Therefore, it is relevant to look more into this mechanism in a more general setting, focusing on how much SES differences in cognitive development are explained by parenting, which are far larger than those presumably generated by family instability or divorce, making its analysis clearer.

## **Theoretical Background**

Social stratification research has identified various sociodemographic correlates of poverty and socioeconomic disadvantage (Brady 2019), which have a substantial impact on family life. On one hand, behavioral theories of poverty make the case that the behavior of individuals, and the families they build, are the main explanatory factors behind the prevalence of poverty and the intergenerational transmission of inequality (Edin and Kefalas 2005; McLanahan 2004). Given that it is about choices made – the micro-behavior – and not much about the circumstances surrounding and

informing those choices, attempting to change or modify harmful behaviors, and in particular behaviors around marriage and family stability, is considered as a potential target for poverty alleviation (Cahill 2005), supposedly largely benefiting the children of impoverished families. The origin of this perspective can be traced back to structuralist-functionalist views of family roles in society (Kingsbury and Scanzoni 2009; Burgess 1950), which in the US context were particularly targeted at single Black women (Moynihan 1965; Wilson 1987), primary breadwinners and caregivers in their families (Nieuwenhuis and Maldonado 2018).

G. S. Becker (1991) was among the first to synthesize and formalize the ‘stylized facts’ known at the time about family behavior and its relation to social inequality. Marriage, divorce, fertility, and stepfamilies, under the individualist choice perspective advocated by him, became the focus of economics and were considered as the outcome of rational costs and benefits comparisons (Browning, Chiappori, and Weiss 2014). Though the topic of family behavior was not new to the field of home economics (Hara 2016), the innovation of Becker came from the application of standard micro-economic models to household decision-making. The rational choice approach, however, fell short and quickly out of flavor within family sociology – and in particular feminist economics perspectives – mainly because of its rather strong assumptions on individual behavior and the lack of empirical support for some of its propositions (Albelda, Himmelweit, and Humphries 2004; Hara 2016). For example, these models were static – in order to be mathematically traceable – and did not address how family behavior could be the consequence of previous events in the family life course Furstenberg (2016).

On the other hand, and in contrast to these perspectives, the studies of Engels and Morgan (1884), and much later those of Bourdieu (1993), both from a social conflict paradigm closer to structural and political theories of

poverty (Brady 2019), considered family the behavior as the result of larger sociological, contextual factors in which family life unfolds, and in which family choices are made. However, for this theory, the determinants of a specific family behavior are difficult to trace. These determinants can include aspects related to family experiences in the parental home, employment trajectories, housing, social policies, neighborhood effects, etc. For example, Elder (2018)'s study showed that the overall historical period in which certain choices were made, and the condition of interlinked lives in which those events are experienced, also turned out to be of substantial importance for a person's life chances as adults. The life course perspective further complicates our understanding of family behavior seen as a dynamic type of behavior (L. Bernardi, Huinink, and Settersten Jr 2019). Research suggests that disadvantages experienced in early childhood can have cumulative and non-linear effects in the future (DiPrete and Eirich 2006). Moreover, recent studies suggest that the family life course has rather small repercussions on poverty and employment trajectories (Fasang and Aisenbrey 2021), which suggest that is economic interventions, and not family interventions, which can bring about socioeconomic improvements in the lives of disadvantaged families. Instead of attempting to change the demographic behavior of poor families to rip the benefits associated with normative family transitions (Hadfield, Ungar, and Nixon 2018), social policies seeking to break down cycles of intergenerational transmission of disadvantage should focus on addressing the large, growing inequities in family resources that affect children's wellbeing and family formation (James J. Heckman 2011). These mechanisms do not operate on parents' choices but do so through the conditions, contexts, and environments in which parents raise their children.

In trying to understand the causes of social inequality, and explore potential policy solutions to its pernicious effects, the behavioral theories of poverty

have provided important insights for the study of family behaviors. However, they have not given enough attention to the problem of endogeneity in the design of public policies aimed at modifying such behaviors. Some attention has been placed on selection bias (Ginther and Pollak 2004; Manski 1993), but in general, endogeneity takes multiple forms that have been only partly discussed, which I explore in-depth in this dissertation. Endogeneity, which refers to the situation when a regressor or an ‘independent’ variable is correlated with the error term (Cameron and Trivedi 2005, 92), occurs when not all causal interdependencies have been appropriately adjusted for in a statistical model that is aimed at explaining the effect of a specific factor or exposure. In the case of family stability, other alternative factors simultaneously affect the probabilities of divorce, repartnering, and children’s wellbeing. For example, it is known that housing and employment policies affect family behavior, but policies that deregulate housing markets and labor markets are likely negatively impacting family stability and parenting in ways that are difficult to observe (Lauster 2010; Jacoby et al. 2017; Desmond and Perkins 2016), and also affecting children’s wellbeing, but not through family behavior. Before trying to influence family behavior through the imposition of direct, or indirect, incentives or penalties, we ought to consider that divorce, parenting, and family instability, in general, are determined by complex selection processes involving cumulative disadvantage, life course effects, and contextual effects. Family behavior is the result of an overly complex bundle of causal interdependencies that highlight the complexity of accounting for endogeneity.

One way of summarizing the contribution of this dissertation is as a special attention devoted to those causal interdependencies that, although implied by theoretical claims, have received little or no attention in the literature. The concern here is not, however, of a purely methodological kind - solutions to the general problem of endogeneity in observational studies

already exist (Yao et al. 2021). Instead, my concern is that these rich theories often imply strong forms of endogeneity that go often unnoticed or are ignored in further theorizing and empirical work. I argue that, instead, these endogenous forms constitute a “richness of sorts” for the empirical study of social inequality. Life course theories suggest that effects of events occurring in the distant past may have long-lasting repercussions in the present, not necessarily following the Markovian property which only considers the recent past. Family instability claims that many consequences in the future follow specific family transitions, but those family transitions are also the result of previous events and affect further family transitions. Considering parenting as a time-varying exposure experienced by children implies, as well, that we ought to think how parenting is affected by SES and what its main confounding mechanisms are. Finally, psychological constructs related to child development are based on strong assumptions about what standardized assessments for the measurement of, for example, cognitive constructs, can actually capture.

Various forms of bias resulting from improper comparisons have affected the analysis of the effects of family behavior on children’s well-being. For example, children exposed to multiple family transitions not only differ from children unexposed to any transition (i.e., those remaining in stable two-parent families) in some background characteristics but also in multiple other time-varying characteristics that are affected by family transitions and that affect family transitions in dynamic, complex ways. Moreover, even when one focuses on the case of a single event taking place in the lives of children, such as the departure of the biological father, the trajectories of factors or confounders that lead to that event are not comparable across children. Another case in point is the study of parenting as a mediating mechanism of the SES gaps in children’s cognitive development, which has neglected the interactions between the different parenting mediating

activities, and overlooked the role of confounding mechanisms of the different parenting mechanisms. In fact, SES also affects many of those confounding mechanisms, which complicates mediation studies in considerable ways, as I explore in this dissertation. Finally, the measurement of cognitive and “non-cognitive” skills in childhood is complicated by the fact that those variables are unobservable and can only be measured with error. Measurement error on the dependent variable, often ignored or assumed not to be a major source of bias, may turn out to be highly problematic for the case of unobservable variables that capture the signal and the noise, noise that is not randomly distributed. These and other forms of improper comparisons are the subject of this dissertation. One could say that the overarching theme in the work here presented is that of the endogeneity of family behavior and its implications for establishing causal associations in family demography. I elaborate on these aspects and explain its relevance for future empirical work in family demography and social inequality more generally.

Instead of focusing on a single theoretical model of demographic behavior, I take a data-driven approach to the study of family behavior and children’s wellbeing. I ask what processes could generate data that may look like the observed children’s wellbeing measures, and what processes generate the associations between these measures and three types of family behavior. Answering this question is what may allow us to rethink which causal mechanisms might be at play. What I mean by a “data-driven approach” can be summarized under one overarching question: What are the data-generating processes of children’s outcomes affected by family instability or parenting? Data on children’s wellbeing are generated by multiple mechanisms. Some of these mechanisms may be related to family instability or changes in family structure, as well as parenting, whereas others, such as neighborhoods, schools, peers, etc., are not. The majority of

the mechanisms generating the data on children’s cognitive or behavioral outcomes go beyond the family, but to understand the influence of one specific factor, such as family behavior, it is necessary to adjust for the alternative pathways that affect or are affected by family instability.

Regarding methodology, this dissertation takes a decided “causal inference” turn in the reading of the main family demography theoretical perspectives (Hernan and Robbins 2020; Pearl 2009). A causal inference perspective takes as a starting point knowledge and assumptions about potential data-generating processes at play to later ask counterfactual type of questions (Rubin 1974; Holland 1986). What would the wellbeing of children of divorced parents or exposed to family instability had their parents not separated? Would differences in language skills among pre-school children be smaller if parents from different SES had the same type of parenting, the same level of parental investments, and the same frequency of parenting practices as high SES parents do? To answer such questions, or to approximate an answer to them, we need to move beyond the “simplifying assumptions” of most of current quantitative social science. Though there is no causal inference in observational studies without additional assumptions about the data generating mechanism, making these assumptions clear is fundamental for proper studies. Much clarity can thus be gained by at least specifying what those data generating processes are.

I attempt to adjust for various selection mechanisms by considering the following general statements. First, family behavior is dynamic, differs over time and by social groups, and it affects and is affected by multiple other socioeconomic factors that are also changing and interdependent. Second, the effects of experiences family instability during childhood may appear much later when children are teenagers or even by the time they reach adulthood. And third, these effects may be nonlinear, something which the main theoretical tenets often remind us of when they discuss the

multiple-way interactions in which effects are produced. To “adjust for” the confounding factors that affect selection into specific family structures or instability trajectories, we should go beyond adding variables in a linear regression framework. If the effects are non-linear, cumulative, or if there are multiple-way interactions, or even if the constructs are not measured correctly and only approximately and with error, then the phrases “net of” or “after controlling for” a given set of -often limited- confounders should be read with caution. First, because there is no actual control being made for these covariates, as in randomized *control* trials; and, second, because the statistical adjustment can be done in multiple ways, and more often than not researchers only show a final adjustment, discarding, without exploring, other potential forms of adjustment bias-reducing strategies.

## Summary of chapters

This dissertation is composed of six chapters, starting with this introduction. Chapters 2 through 5 are single case studies that examine one aspect of the negative effects of family instability and parenting on children’s wellbeing. These are all single-authored papers that were written to be published in peer-review Journals. Final chapter 6 presents the overarching conclusions and a general discussion, as well as future work.

Chapter 2, *Fair comparisons: life course selection bias and the effect of father absence on US children*, published in **Advances in Life Course Research, Vol. 51 (2022)**, starts from the premise that father absence in opposite-gender couples has been shown to have detrimental effects on various measures of children’s wellbeing encompassing health, behavior problems, and cognitive development, net of selection bias (McLanahan, Tach, and Schneider 2013). However, though often made, such a claim has never accounted for the trajectories of confounders, such as housing, employment, or health. Life course informed research suggests that

adjusting for selection bias, even when only looking at specific transitions, may be more complex than what has been thought so far. In this chapter, I show how important it is to adjust for the trajectory of confounder covariates in a nonparametric fashion – meaning without making assumptions about how the effects are mostly additive and linear – when estimating how father absence affects children’s wellbeing. I use data from the Fragile Families and Child Wellbeing Study (FFCWS 2019) to estimate the total effect of the departure of the biological father on children’s wellbeing, considered as a single time point transition, and I further estimate the delayed or fade-out effects of this family transition. I employ Bayesian additive regression trees (BART), which is a machine learning and causal inference method suited for statistical models that involve a number of covariates (J. L. Hill 2011). Because I am adjusting for the trajectories of these confounders, the method is especially well-suited to the study of these effects on adolescents. The main result of this chapter is that, after adjusting for multiple time-invariant and -varying confounder covariates, as well as their history, the obtained estimates of father absence’s effect on children’s wellbeing are substantially reduced. I refer to this finding as life course selection bias because it captures the idea that the trajectories also increase selection for divorce – and hypothetically for other family transitions too – in a world that does not follow the Markov property. The results suggest that early and middle childhood are not negatively affected by the departure of the biological father in any of the dimensions I look at. However, life course selection bias mostly affects the estimates of father absence on adolescence. I believe this relates to the fact that children directly experience their parent’s confounder trajectories leading to divorce or separation, which does not occur when father absence is experienced in early childhood. The main conclusion of this paper is that father absence is mostly a marker of life course cumulative socioeconomic disadvantage, and

not a cause of negative effects.

In Chapter 3, *The complex effects of family instability on adolescent problem behavior in a US birth cohort*, currently under review in the journal

**Demography**, I turned my attention to family instability. This concept refers to the number of family transitions experienced by children, including the departure of the biological father and also the entrance or exit of a social father. Family instability has also been hypothesized to negatively affect children's wellbeing, even after accounting for confounding factors or selection bias (Lee and McLanahan 2015). However, once again, a life-course reading of the family instability hypothesis reveals crucial interlinkages between family life and employment, income, household, and housing trajectories, which explain away part of this negative impact. Time-varying confounders affected by previous episodes of family instability, and affecting future family stability, can generate what is known in epidemiological literature as treatment-confounder feedback bias (Hernán and Robins 2006). This bias takes place in dynamic settings and for time-varying exposures, when the exposure episodes can affect confounders intertemporally. Despite being a form of dynamic or time-dependent confounding that occurs rather frequently in sociological research, it has not received much attention in sociological research until recently. Again, in this chapter I employ data from the Fragile Families and Child wellbeing Study to empirically show this on one dimension of children's wellbeing: behavior problems. Problem behavior, including both externalizing and internalizing problems, is one particular dimension of children's wellbeing where effects of family instability are often found in the literature (Cavanagh and Huston 2006; Fomby and Osborne 2017). Here the effects of interest, which correspond to a life course cumulative instability model and a life course pathways model – distinguishing the effect of timing of exposure to family changes as well as the number of changes – are obtained through doubly robust marginal

structural models and inverse probability of treatment weighting (Hernán, Brumback, and Robins 2000). I further explore the potential for heterogeneous effects by the child's gender assigned at birth and their racial-ethnic background. The results indicate that a dynamic version of the selection hypothesis should be considered as an alternative explanation to the instability hypothesis, one that better explains the differences in children's problem behavior when they are exposed to family instability. Effects of family instability on problem behavior are small and often not statistically significant, and their size is reduced after accounting for a set of biasing feedback mechanisms. Although it seems that three changes in family structure have the largest, though not statistically significant, negative impact, the effect of four and up to five changes have a smaller effect, which goes against the family instability hypothesis. Boys and White children seem more affected by a high number of changes than other demographic groups. The main conclusion of this chapter is that children's wellbeing is equally affected by the reinforcing and counteracting processes brought about by family instability, which may lead us to reconsider the interlinked pathways between trajectories in the socioeconomic context and family life.

In Chapter 4, *What can parents do? The causal mediating role of parenting in SES differences in children's language development*, currently under review in the **Journal of Family Research**, again well established social inequalities constitute the starting point. Research has shown there are large differences in children's language development by parental SES (Madigan et al. 2019). Various authors have argued that these differences are the result of parenting behavior (Greg J. Duncan, ZiolGuest, and Kalil 2010; Fomby and Musick 2018). SES gaps in language skills among preschoolers, these authors argued, could be substantially reduced by intervening on parenting styles, practices, and parental investments (Ayoub, Vallotton, and

Mastergeorge 2011; James J. Heckman, Humphries, and Kautz 2014; Price 2010). However, something that is still unknown, from a causal inference perspective, is the extent to which parenting causally mediates the effects on language skills of growing up in low-SES contexts. In this chapter, I employ data from the National Educational Panel Study starting cohort 1 (Blossfeld, Roßbach, and Maurice 2011), which is a random sample of  $N = 1892$  children that were born between 2012 and 2013 in Germany. In terms of methodological choice, I employ interventional causal mediation analysis to estimate the mediated share of the total effect of SES on children’s language that goes through parenting (Nguyen, Schmid, and Stuart 2021; T. J. VanderWeele and Tchetgen Tchetgen 2017). The results of these analyses suggest that joint parenting explains around one third of the total effect of SES on early language skills, but close to nothing of later language skills. These mediated shares are remarkably small given the overemphasis made on intervening on parenting as a potential solution. Although an important share of the SES effect operates through this specific type of demographic behavior, and parenting practices do affect children’s early language skills, hypothetical interventions on the parenting of low SES parents would have a limited effect on closing the language skill gaps in this cohort, and especially limited when considering later gaps observed when children are older. These results can be explained by the multiple alternative pathways through which inequality in language skills operates, pathways that do not involve parenting, and that may explain a larger share of the SES differences.

Finally, in Chapter 5, *Is it just noise? Measuring unobservable cognitive abilities in early childhood*, published in **Personality and Individual Differences, 166 (2020)**, the starting point is the large evidence suggesting that children from disadvantaged backgrounds have lower cognitive abilities than their socioeconomically advantaged peers (Noble et al. 2015; Greg J. Duncan, ZiolGuest, and Kalil 2010; Farah 2017). However,

there are a few considerations worth the attention of family demography scholars before one can finally establish with certainty the size of associations between family demographic behavior and children's cognitive development. Two characteristics of the measurement of cognitive constructs, as understood in mainstream psychometrics (Association, Association, and Measurement in Education 2014), make it difficult to quantify these inequalities. First, Bond and Lang (2013) suggested that although items used within a standardized test may provide ordinal information allowing us to rank children, these tests do not conform to the properties of an interval scale that is necessary to quantify inequalities in accordance to scientific principles of measurement. Second, and more problematic for this quantification, is that a causal understanding of validity is incompatible with the standard validation framework applied in psychology (Borsboom, Mellenbergh, and Heerden 2004, 1067), further complicating what it actually means to measure unobservable constructs related to cognitive development. These two problems, but in particular the second one, represent large, though unexplored problems for the estimation of causal effects on unobservable variables in the context of family research, as well as other types of exposures or treatments. The reason for this is that measurement invariance (Penfield and Camilli 2006; Uher 2020), an indicator that the measurement of these unobservable constructs does not function as intended, generates endogeneity bias in causal inference for the estimation of the effect of variables of interest on test scores, and may make comparison across groups invalid (Kuroki and Pearl 2014). In this paper, I analyze these data problems using three standardized assessments taken by German children, thus taking an empirical look at the problem of measurement error. I explore the limits of standardized assessments by employing nonparametric psychometric models and the representational theory of measurement. An alternative framework for validation of such

constructs for use in demographic research, and other fields, built on these two methods, may help to determine whether data fit the assumptions of a measurement model, and to overcome some limitations when it does not. I further compare competing statistical modeling alternatives that reveal substantial differences in the quantification of social inequalities. The main conclusion of this chapter is that, measurement error does not behave like simple random noise, and, after accounting for it through this alternative validation framework, I find an unsettling reduction in estimated effect sizes associated with factors often studied in social stratification.

One common feature among the different chapters is the focus on understanding the determinants of child development and children's wellbeing. What factors explain the differences seen in the behavioral and developmental pathways of children? Previous research considers childhood family experiences, encompassing family structure, family instability and parenting, as important determinants in a child's life, with substantial consequences in adulthood, mainly because the family is the first environment in which children spent most of their time, only matched perhaps with schools and peer-networks. However, major determinants of such key family experiences, that also affect children's behavior and developmental pathways, are parental socioeconomic status and various parental socio-demographic characteristics that vary over time. Moreover, it is worth emphasizing that for children from opposite-gender parents both the mother and father side characteristics matter. The different chapters in this dissertation attend to both of these factors by considering them as interrelated, though unique contributors to child development, something that most previous research does not attend to. In causal inference language, mother's and father's characteristics are distinct confounders of the association between family behavior and children's wellbeing. Researchers should therefore attempt to capture all parental characteristics

when estimating the effects of any given family behavior of interest on children's well being and child development more generally.

The focus on causal inference has implications as well for the type of methodologies necessary to answer specific causal queries about the effects of family instability. Only some methods are useful to disentangle the causes of effects and effects of causes, in particular when the exposures of interests are time-varying, in contrast to single time point events. In this dissertation, I make use of various methodologies developed within epidemiology but which have important advantages with respect to more traditional methods used in sociological research (e.g., generalized regression, matching, instrumental variables, etc.), and are discussed in the following.

Although the different chapters have a methodological focus, they are based on data coming from two specific countries: the US and Germany. None of the chapters compares the US to Germany, in fact, none of the chapters has a country comparison focus. However, it is worth highlighting that these countries differ markedly in various family life domains and, therefore, the effects of family behavior on children will depend on the specific contexts where one examines them. For example, the prevalence of divorce is considerably higher in the US than in Germany, and family instability is much more common in the US. More importantly, it is likely that the selection processes affecting the probability of divorce or repartnering, and therefore the effects of these two family transitions, as well as others, on children, are markedly different. The US is an extreme case where the effects ought to perhaps be the largest. Whereas in Germany, partly thanks to the socioeconomic welfare-state poverty alleviation programs and the supply of public services such as daycare and schools, the effects of family transitions and specific types of parenting ought to be rather small. Therefore, in these two countries we ought to expect highly different interdependencies among life domains. Future research, with a more

country-comparative perspective, could help in further elucidating whether effects vary in the hypothesized directions.

## CHAPTER 2 - Fair comparisons: life course selection bias and the effect of father absence on US children

*This paper has been published as: Rodríguez Sánchez, A. (2022). Fair comparisons: Life course selection bias and the effect of father absence on US children. Advances in Life Course Research, 51, 100460.*

*This is the accepted manuscript. The final version is available at <https://doi.org/10.1016/j.alcr.2021.100460>*

Father's departure from the household in opposite-gender couples because of separation or divorce is a presumed cause of negative effects on children's wellbeing. Although father absence thus defined does not rule out the possibility that fathers may still be in contact with their children, the literature on father absence considers that the very fact of children not being co-residents with their fathers implies a loss of resources and time, less effective parenting, potential inter-parental conflict, and little contact with the father (McLanahan, Tach, and Schneider 2013), which are not compensated by other custodial arrangements or stepfathers. Studies have shown that children's exposure to episodes of father absence is often associated with worse outcomes for children (Fomby and Cherlin 2007). It is further argued that the negative effects of father absence contribute to the intergenerational transmission of inequality and disadvantage (McLanahan 2004; McLanahan and Percheski 2008). However, if the departure of the biological father out of the family unit experienced during childhood can have such long-lasting negative consequences on children's lives, research in family sociology should also consider the long-lasting effects of other events across the life course which may trigger divorce or separation in the first place, and which may also affect children's wellbeing.

In other words, the trajectories leading to divorce or separation may also be relevant when comparing children exposed to father absence and children in stable families. One consequence of this is that family formation and dissolution processes may not follow the Markov property, which states that change in a given process (e.g., family instability) depends on the present state and does not depend on its history. However, if events occurring in the distance past still linger and have cumulative or non-linearly effects in the future (DiPrete and Eirich 2006), then an adjustment for trajectories leading to divorce is crucial for a fair comparison between children's wellbeing in two-parent and single-parent families. A fair comparison between children who experienced a family transition, such as the departure of the father, versus those who did not, but could have experienced it, should account for the role of the socioeconomic trajectories, such as employment trajectories, that may lead to divorce or separation (e.g., earlier educational and employment trajectories, or housing instability, as well as a downward income, occupational or residential trajectories, or even a deteriorating relationship quality).

Although family instability scholars make an analogous claim about the long-lasting effects of childhood family instability (Amato and Patterson 2017; McLanahan 2004), selection bias has not been considered from this perspective, despite studies on intergenerational transmission of divorce suggesting more complex dynamics leading to father absence (Amato and Patterson 2017). Life course studies have argued for the strong linkages between events that occurred in the distant past and those which occur in the present (Elder Jr, Shanahan, and Jennings 2015). Stable and unstable families are likely to differ in their life courses up to the time point when a family transition takes place, and, therefore, the history of confounder covariates, and their interactions, should be accounted for to disentangle long sequences of causality. If family dynamics do not follow the

Markov-property, this could imply research on its effects on children’s wellbeing should account for more complex forms of selection. Although previous studies have argued for addressing omitted variable bias (McLanahan, Tach, and Schneider 2013), only adjusting for the recent past and time-fixed characteristics leaves open still the possibility for selection bias in the estimation father absence effects.

The first aim of this paper is to advance life course research and family demographic research on cumulative disadvantage (DiPrete and Eirich 2006; L. Bernardi, Huinink, and Settersten Jr 2019), which suggest more complex mechanisms may lead to divorce or separation. Second, the paper employs a methodological approach that combines machine learning and causal inference to get precise estimates of the causal effect of the departure of the father on various measures of children’s wellbeing adjusting for selection based on the trajectories of confounders. I develop the concept of *life course selection bias* (i.e., the selection of the trajectory of confounders), and methodologically I implement an adjustment for this using Bayesian additive regression trees (BART, Tan and Roy 2019). This nonparametric method combines regression trees and regularization priors in a Bayesian framework to find a more appropriate balance in the distribution of relevant confounders of the association between father absence and children’s wellbeing. The BART algorithm outperforms both conventional as well as other machine learning alternatives in causal inference tasks (Chipman et al. 2010; Dorie et al. 2019). BART is based on a data-driven approach to model the response surface – measures of children’s wellbeing – jointly and flexibly and the treatment assignment mechanism (i.e., the exposure to the departure of the biological father) in a nonparametric way, which is useful when little knowledge exists around the correct model specification for the outcomes of interest, and the number of potential confounders increases.

## Background

### **Father absence in US family sociology: Which causal effects?**

The interest in father absence began as researchers noticed the changes childbearing and family formation patterns that have been taking place over the past 70 years in the United States (Furstenberg 2016), which decreased intergenerational economic mobility because of the loss of material resources and parenting time experienced by children (McLanahan and Percheski 2008). However, the sociological father absence literature presents two main hypotheses around this issue: one in which father absence has negative effects on children's wellbeing, leading children from low resource families through a divergent path than their advantaged peers (McLanahan 2004), and another one where selection bias into this type of family transition explains away most if not all the effects of father absence (McLanahan, Tach, and Schneider 2013).

The study of family instability has a large and strong tradition in sociological research, as over a dozen literature reviews attest to (Amato and Keith 1991; Amato 2001, 2010; F. Bernardi et al. 2013; Esping-Andersen 2016; Hadfield et al. 2018; Härkönen, Bernardi, and Boertien 2017; Haveman and Wolfe 1995; Jeynes 2006; McLanahan, Tach, and Schneider 2013; McWayne et al. 2013; Raley and Sweeney 2020; Saint-Jacques et al. 2017; and Wells and Rankin 1991). Most up-to-date systematic reviews on the topic emphasize that evidence for the family instability hypothesis is mixed (Hadfield et al. 2018; and McLanahan, Tach, and Schneider 2013). Although some studies suggest that the association is not causal, or that selection bias explained it away (Bhrolchain 2001; Erola and Jalovaara 2017), other studies consider that more robust evidence still suggests smaller negative effects remain on some dimensions of wellbeing (Lee and McLanahan 2015). Despite efforts to reconcile the father absence

and the selection hypotheses, as seen in the claim that negative effects of father absence exist but are small and therefore not so substantial for most children (Amato 2003; McLanahan, Tach, and Schneider 2013), underlying this still debated question are at play two different causal understandings or models of family dynamics (Raley and Sweeney 2020).

On one hand, the hypothesized negative effects follow from a deficit model based on structural-functionalists premises (Kingsbury and Scanzoni 2009, 307), where father absence results in an irreplaceable loss of resources deemed fundamental for child development, a model which is normatively charged and considered outdated (Hadfield, Ungar, and Nixon 2018; Coltrane and Adams 2003). The departure of the father makes up a source of distress in multiple domains of a child's life, although research has also discussed the benefits of separation or divorce for some families, especially those in which women live in stressful or violent relationships (Kelly 2000; Fox et al. 2002; Hetherington and Stanley-Hagan 1999). Father absence may affect the psychological and social functioning of children, especially during early childhood (Cavanagh and Huston 2008). Although children will experience other transitions in their life course (Fomby and Mollborn 2017), the distinctiveness of the effects of father absence results from the substantive role that fathers play in shaping children's life course and from this transition being often the first one.

Generally stated, the family instability hypothesis claims that changes in family structure bring about disruptions in the lives of children (Cavanagh and Fomby 2019), which require adaptation and adjustment from all family members, and which in turn may generate stress on children and their mothers (Masarik and Conger 2017). The departure of the biological father out of a married or unmarried two opposite-gender parent family unit disrupts the family routine and reduces the time, social capital, and financial resources of the household (Fomby and Cherlin 2007; Härkönen,

Bernardi, and Boertien 2017), and corresponds to the first and main negative shock in a trajectory of family instability. As reviewed in McLanahan, Tach, and Schneider (2013), most studies are focused on the effect of father absence. Father absence brings about stressful and negative events in the life-course of adults and their children because children's parents cannot fully benefit from the gains of marriage (e.g., G. S. Becker 1973; and Schultz 1974; more recently Browning, Chiappori, and Weiss 2014), regardless of whether children stay in close contact with their fathers after divorce or separation. For example, single parent families have fewer resources than nuclear families; and blended, complex or stepparent families provide less than optimal investments in children (Browning, Chiappori, and Weiss 2014, 438–70). It is further argued that not having a father present in the household, even when fathers remain in contact with their children, also results in negative effects on children who lack the influence of positive masculine role models in their development (Sigle-Rushton and McLanahan 2004). Mediated by its effects on the psychological wellbeing of the mother, separation or divorce may negatively affect the quality of parenting that the resident mother provides to her child (Fomby and Osborne 2017). However, Cavanagh and Fomby (2019)'s remark on the lack of evidence for the mediating role of many of the assumed pathways leading from father absence to negative outcomes seems to suggest that the association between father absence and worsening children's outcomes still has not completely resolved the problem posed by endogenous family behavior (Manski 1993).

On the other hand, scholars have also stressed that the sources or causes of father absence are at the center issues of socioeconomic inequality. They consider father absence as a correlate of disadvantage rather than a cause of it. The two opposite-gender parent family, also referred to as the *Bourgeois* family in Engels and Morgan (1884), has always depended on a set of often neglected historical and sociological premises that make a certain family life

possible for some, but unattainable for others (Coontz 2016; Bourdieu 1993). For example, low income single-mothers are often unable and not unwilling to marry because they lack the means to do so (Edin and Kefalas 2005), suggesting that it is the conditions for a proper marriage (e.g., a proper education, a stable job, and secure housing) which have become more difficult to attain for them and their prospective partners, rather than an abandonment of the marriage norm as the functionalist premises lead us to think (Cherlin 2004). It is these same factors –socioeconomic status, neighborhoods, economic insecurity, etc.– which affect child development (Minh et al. 2017; Conrad-Hiebner and Byram 2020; Pace et al. 2017; Devenish, Hooley, and Mellor 2017). Therefore, it is socioeconomic factors which explain divorce or separation and affect children’s wellbeing.

Although good theoretical reasons exist for why instability in the romantic relationship history of parents may negatively affect children, exposure to father absence and family instability are not experienced by all families at similar rates (Cohen and Pepin 2018). The propensity of individuals to experience divorce or separation differs in a population and may generate selection bias for estimating the effect of father absence on children’s wellbeing (Hadfield et al. 2018; Härkönen, Bernardi, and Boertien 2017). Research in this area stresses the role of selection due to confounding factors, those which affect the risk of experiencing father absence and children’s well-being. The selection hypothesis proposes that factors such as low income, unemployment, or low education, which are associated with changes in family structure, drive the negative association found in the data (Jackson 2016). Critics of the father absence literature argue that the negative association between the departure of the biological father and children’s well-being follows from those parental characteristics that make children from specific backgrounds more prone to experience disruptions in the partnership trajectories of their parents.

### **Life course selection bias**

The selection hypothesis, as selection bias is known, is a competing *causal* hypothesis that explains the small negative effects of father absence by considering what happens before, and therefore *causes* the departure of the biological father. However, this hypothesis has not been considered from a life course perspective. Employment, educational, housing, living arrangements, health, relationship quality, and wellbeing *trajectories* of parents, which may lead to divorce or separation, may further confound the negative association between the departure of father and children’s wellbeing.

It is worth noticing that the prevalence of worsening trajectories can vary by racial-ethnic groups, as defined by overlapping categories of gender and race-ethnicity (Lee and McLanahan 2015; Cavanagh and Fomby 2019; Cohen and Pepin 2018), as suggested by intersectional research (Collins 2009, 84). The thinking around confounder trajectories and their effect on family structure was implicit in the pioneering works of Du Bois (1899) and Frazier (1928), when these authors discussed the higher prevalence of single motherhood among US Blacks. Both authors pointed out that the root causes of problems affecting Blacks were traced to the effects of racial oppression and poverty, especially as they related to inadequate housing and segregation, Blacks’ low educational levels, and their low occupational class attainment, and not on the absence of any family structure.

In other words, what may explain that children exposed to father absence tend to, on average, do worse on various measures of wellbeing than children raised by their two biological parents, are the negative and self-reinforcing *cumulative experiences* that build up and lead parents to divorce or separate, and that affect children’s wellbeing, regardless of their racial-ethnic background. These negative experiences lead parents who follow a particular life course to be exposed to transitions in their partnership trajectories with

more frequency than others, whose life course sets them apart from other pathways in their family formation history (Aisenbrey and Fasang 2017).

I argue that estimating the total effect of a family transition such as the departure of the biological father out of the family unit should also account for *life course selection bias*. Following a life course perspective, this concept captures the specific biasing effect that the confounders' trajectories may have in effect estimation dynamically. Dynamic confounding is fundamentally linked to the departure of the biological father, although is often considered in more complex forms of selection as studied when examining the effects of number and type of family transitions experienced during childhood (Lee and McLanahan 2015; Wu 1996). However, father absence corresponds to the first of the many other family transitions children might experience, perhaps the most critical one.

*Life course selection bias* corresponds, thus, to the effect that trajectories of confounders, their history, have on the risk of divorce or separation. This goes beyond the simplification that only the immediate past matters and extends our consideration of potential confounding to other multiple-ways in which confounders' trajectories or history interact. For example, Hansen (2005) shows that unemployment episodes increase divorce or separation risk (Wagner 2020), where parents unemployment also negatively impacts children's wellbeing (Nomaguchi and Milkie 2020). Not only one episode of unemployment generates negative effects, but recurrent unemployment episodes may have similar and cumulative effects as well, and as some argue is a difficulty for single mothers in urban US (Edin and Kefalas 2005). It is in this sense that the probability of divorce as a function of unemployment may not correspond to a memory-less process. Therefore, following principles such as cumulative disadvantage (DiPrete and Eirich 2006), which imply confounders' trajectories would matter because of their effects on children's wellbeing and on the risk of exposure to father absence,

adjustment for confounders' trajectories might be relevant for estimating the effect of father absence on children's wellbeing.

Prior processes leading to family instability, as seen in the trajectory of confounder covariates, are as important in both the selection of individuals into specific family structures and the wellbeing of children (Amato 2010; Cavanagh and Huston 2008; and Morrison and Cherlin 1995). They may even make up the main confounding mechanisms. Life course studies of demographic behavior suggest that biographical, historical, and ecological changes are as relevant, as the immediate past, for unraveling causal relations between events that occur at different time points (Elder Jr, Shanahan, and Jennings 2015; and Macmillan and Copher 2005). However, the role of the not so immediate past has been ignored by earlier research portraying family instability as having long-lasting consequences (McLanahan 2004; McLanahan, Tach, and Schneider 2013).

## Methods and Data

### Causal inference for family transition effects, re-examined

From a causal inference perspective, the departure of the biological father is as a point-in-time event that may occur to a child and her mother at some specific point in their life courses, for most children happening only once in their lifetime (Turney and Halpern-Meekin 2020). A dichotomous variable  $Z \in \{0, 1\}$  denoting the time when the father left the household may represent this event. Following the notation of Holland (1986) and Rubin (1974), the focus of the father absence literature is on comparing potential outcomes on some child wellbeing indicator between children who experienced this form of family instability,  $Y_i(Z = 1) = Y_i(1)$ , and children who did not experience it,  $Y_i(Z_i = 0) = Y_i(0)$ , that is under  $Z = 1$  or  $Z = 0$ , respectively. It is not possible to estimate the difference in potential

outcomes for any given child  $i$ , represented by  $Y_i(1) - Y_i(0)$ , because only one of these potential outcomes is observed:  $Y_i = Y_i(1)Z_i + Y_i(0)(1 - Z_i)$ . Therefore, instead, the alternative is to estimate the sample average treatment effect, denoted by  $\sum_{i=1}^n (Y_i(1) - Y_i(0))$ , and the sample average treatment effect on the treated, denoted by  $\sum_{i=1; z_i=1}^n (Y_i(1) - Y_i(0))$ , or a conditional version of these, the conditional average treatment effect (CATE).

Specifically, this implies that the role of two crucial assumptions of causal inference, unconfoundedness and common support (J. Hill 2011), deserves special reconsideration if indeed *life course selection bias* drives to some extent these associations. Unconfoundedness, the first of these assumptions, denoted by  $Y(0), Y(1) \perp Z | \mathbf{X}$ , means that given a set of covariates  $\mathbf{X}$ , it is safe to assume that children’s potential outcomes are independent of the family transition experienced, where  $\mathbf{X}$  contains all confounder covariates occurring before the father leaves the family unit. Although a solution might be to include the trajectories of all these confounders into statistical linear models to adjust for remaining confounding, such a strategy could be bias increasing because of a small sample size relative to the number of covariates (Middleton et al. 2016). The second assumption is that of common support which says that the probability of experiencing this treatment must be greater than zero for all children,  $0 < Pr(Z = 1 | \mathbf{X}) < 1$ , thus guaranteeing a sufficient overlap between the two groups under comparison. Additionally, as stressed in J. Hill and Su (2013), OLS estimation using a matrix of confounder covariates  $\mathbf{X}$  that is high-dimensional compromises the strong ignorability assumption, as would happen when adjusting for confounders’ trajectories, and especially so when employing parametric statistical models. These issues, when unattended to, may lead to extrapolation and lack of common support (i.e., treated units have no comparable data points for a credible counterfactual to be found).

From this perspective, if indeed is necessary to adjust for *life course selection bias*, estimating the effects of the departure of the biological father must overcome three major difficulties. First, given that is unfeasible to assess the effects of family instability through a controlled experiment, and given that researchers lack a proper natural experiment randomizing allocation to family instability (Corak 2001; Gruber 2004; and McLanahan, Tach, and Schneider 2013), selection-on-unobservables is the main potential source of bias. By assuming selection-on-observables and under a life course perspective, however, current strategies aimed at adjusting for as many confounder covariates as possible employing improved surveys or specific methods that account for time-invariant confounders reach their limit by including more confounder covariates to adjust for selection bias.

Second, lack of common support turns more likely by adjusting for more confounder covariates to tackle omitted variable bias (J. Hill and Su 2013), meaning that is less and less possible to find comparable children on all covariates among exposed and unexposed children. Given that confounder covariates may also play a role dynamically, and not just statically, adjusting for baseline confounder covariates is not enough. Previous research suggests that the biographical elements of interrelated family members (mother, father, and their children), in multiple socioeconomic domains, should also be considered in estimations of the causal effect of father absence in so far as it cumulatively contributes to the probability of experiencing a family transition, and therefore to selection into divorce or separation. To do so, however, would require adjustment for an even larger set of confounder covariates and their potential dynamic and inter-temporal interactions.

Third, and because of these two preliminary considerations, previous research has often relied on untestable and strong parametric assumptions (J. Hill and Su 2013). Statistical inference relies on assuming a correct model specification to estimate the effects of family transitions, and on the model's

ability to extrapolate to areas where no data was observed. This is not satisfied when there are not enough comparable units in the treatment or control groups. Therefore, the question remains whether the effects of family transitions would hold, given that models used to estimate assignment to treatment or to estimate the response surface, or both, are mis-specified, a problem that would give rise to a lack of common support and extrapolation. Methods such as propensity score matching may overcome the reliance on parametric assumptions, but they also rely on specific parametric assumptions used to capture how the treatment assignment takes place. The balancing properties of the propensity score depend on these assignment models also being correctly specified; again, an untestable assumption when treatment assignment models include more confounder covariates. Given that the child’s gender or their racial-ethnic group may moderate the effects of family transitions, simply including these covariates in the model might not be appropriate if effects indeed differ by subgroup. Separate models can be fit to account for this possibility, as in stratification or subgroup analyses, but they inflate the rate of type I error because they would artificially increase the chances of finding a statistically significant effect by the mere act of performing more comparisons, more so if we were to include an interaction between gender and race-ethnicity (Schulz and Grimes 2005).

### **Bayesian additive regression trees**

More flexible methodological approaches are thus needed. Unbiasedly estimating the effect of interest which corresponds to the conditional expectation of the outcome  $Y$  given that children have been exposed to the divorce or separation of their parents  $Z$ , requires adjustment for all observed confounding covariates, as in  $\mathbb{E}(Y|Z, \bar{\mathbf{X}} = \bar{\mathbf{x}})$  where the matrix  $\bar{\mathbf{X}} = (\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_{t-\tau})$  denotes the trajectory of a relatively large multidimensional set of confounder covariates, and also the time-invariant

ones, and where  $\tau$  denotes the period before divorce or separation. For a family transition from father presence to father absence, this would mean accounting for all observed factors that may lead to this transition and that may affect children’s well-being. However, the adjustment set should not include mediating factors that explain, in part, the hypothetical effects of father absence, such as marital or relationship quality, or income and employment of the mother after the father left the household.

For example, a downward income trajectory, precarious working conditions over the working life of the child’s parents, or housing instability, up to the point before divorce or separation, may put more pressure on families and cumulatively affect family stability and children’s wellbeing. To prevent estimating the conditional expectation with further bias, we should look for alternatives to the generalized linear regression models, thus avoiding the curse of dimensionality, poor balance, lack of common support, and a strong reliance on untestable parametric assumptions. These often-neglected forms of bias easily appear in this scenario because a multiplicity of factors explains family dynamics. Moreover, if *life course selection bias* plays a role, as I advance in this paper, the trajectories of the confounder covariates are also of importance, and we should therefore adjust for the confounder covariates’ trajectories.

In this context, machine learning and causal inference may provide a convenient solution to the challenge of estimating the effects of complex family dynamics (Dorie et al. 2019; and Molina and Garip 2019). As discussed in J. Hill and Su (2013), these methods can address more complex selection processes, as well as those other sources of bias so far unexplored in the literature on family instability. One such method corresponds to BART. BART could balance the distribution of the relevant confounders of father absence and children’s outcomes by combining two elements: a sum-of-trees and a regularization prior. Given that estimating the

conditional average treatment effect is equivalent to the evaluation of two response surfaces, each corresponding to the distribution of the outcome variable of interest under the two types of exposure, namely  $\mathbb{E}(Y(1)|\bar{\mathbf{X}}) = \mathbb{E}(Y|\bar{\mathbf{X}} = \bar{\mathbf{x}}, Z = 1) = f(1, \bar{\mathbf{x}})$  and  $\mathbb{E}(Y(0)|\bar{\mathbf{X}}) = \mathbb{E}(Y|\bar{\mathbf{X}} = \bar{\mathbf{x}}, Z = 0) = f(0, \bar{\mathbf{x}})$ , one response surface corresponding to the conditional expectation for children who experienced the departure of the father, and another one for children who remained in stable families. BART can flexibly estimate each of these unknown functions in a nonparametric fashion.

The first element of BART, the sum-of-trees, is a sum of  $J$  binary decision trees which split the sample following the most predictive covariates of the treatment (i.e., father departure of the household) and the outcome (i.e., an indicator of children’s wellbeing). Each tree corresponds to a nonparametric model for the outcome variable  $Y$ , and the algorithm constructs multiple trees and combines them. The second element of BART refers to regularization priors corresponding to the number of trees, the variables on which to split, as well as their values, and other parameters used to fit these trees, where regularization refers here to adjustments to the method used to reduce its generalization error. The role of these regularization priors is to prevent BART from overfitting the model to the observed data (Chipman et al. 2010; and J. Hill 2011). BART’s algorithm employs Markov chain Monte Carlo methods, similar to the logic behind ensemble learning in boosting (Friedman 2002). The joint estimation of all parameters is derived from the posterior predictive distribution of  $Y$ . Thus, BART is a desirable alternative to more traditional approaches based on generalized linear models because it may help us overcome some of the strong assumptions on which the father absence literature, as well as sociological theory, rest (Abbott 1988). J. Hill, Linero, and Murray (2020) summarizes the method and its underlying assumptions, as well as computational details. A more detailed explanation

of the methodological aspects of the algorithm is provided in **Supplementary Materials - Chapter 2**.

## **Data**

I estimated the effects of the departure of the biological father using data from the Fragile Families and Child Wellbeing Study (FFCWS). In this study, researchers have followed for over twenty years a cohort of American children that were born between the years 1998 and 2000 in US cities with a population greater than 200,000 (Reichman et al. 2001). The study is based on a probabilistic sample with a complex design in which cities, hospitals, and beds in the selected hospitals were randomly and sequentially sampled, to arrive at a final sample of births in large US cities. For an overview of the data response rates, sample weights and sampling designs see FFCWS (2019) and Kennedy and Gelman (2018). The FFCWS over-sampled births to unmarried opposite-gender parents, which experience more family instability, but it also captured many married and unmarried couples who were living together at the time of the birth of the child (approximately  $n \approx 2000$ ). Although the sample design sought to capture disadvantaged families, children from more advantaged backgrounds were, however, also part of the sample. These advantaged children are, however, the least exposed to father absence or family instability (Kalmijn and Leopold 2020), and the least affected by the hypothetical effects (Cavanagh and Fomby 2019). Table 1 provides an overview of all variables used in the analysis, as well as their construction or operationalization, and the original questions on which they are based.

## **Variables: treatment, confounders, and outcomes**

The treatment or exposure of interest is the departure of the biological father. I compare children who were still living together with both of their

biological parents in one wave, to children who stopped being coresident with their father and lived only with their mother in the next wave, but excluding the few cases where this was because of father's death. Here I focus on the first departure of the biological father, the only one for most children, and the most important and consequential one for later changes in family structure. As shown in , with each wave the analytical sample reduces to the remaining stable families. Children in those families that did not split between waves correspond to the "control" group, a group that shrinks over time, whereas the families in which the biological father left are the corresponding treated or exposed group. In these analyses, I consider the first 15 years of follow-up. At each wave, family structure was coded as being: a) biological two-parent married; b) biological two-parent unmarried; c) divorced or separated, but the mother with another partner; and d) divorced or separated, but with the mother remaining single. Only cases where the child lived with the mother were considered, since the group of children living with the father after divorce or separation is considerably small.

I coded the departure of the biological father as a binary indicator for children who were born to opposite-gender two-parent married or unmarried couples, but who were not coresident with their biological father anymore at the next wave. This is a necessary abstraction given the overall focus on father absence, though it does not allow for a further examination of different custodial arrangements after the father's departure, which may moderate the effect. Given that the entrance of a social father often follows the departure of the biological father, in order not to mix up the effect of father absence with the entrance of the stepfather, I did not include cases where both transitions occurred between waves. This makes possible distinguishing one effect from the other. After the departure of the biological father took place, children exposed to this transition were

excluded from the analytical sample for the estimation of effects in subsequent waves to only capture the timing of the effect. Therefore, the analytical sample at time  $t$  is composed of children in remaining stable families in the previous wave. I did not consider the cases in which both biological parents started a cohabitation spell after the child's birth because they did not make up stable families from birth onward.

To provide leverage on the selection-on-observables assumption, I selected two large sets of potential confounder covariates: a time-invariant and a time-varying one. All models adjust for the following time-invariant confounders: child's gender assigned at birth; low birth weight (yes or no); whether mother consumed alcohol, tobacco or other drugs during pregnancy (yes or no indicators); mother and father's ages at time of birth; their educational levels (i.e., less than high-school; high-school or GED; some college or technical education; and college or graduate education); their self-assigned racial-ethnic categories (White, non-Hispanic; Black, non-Hispanic; Hispanic; and other); their religiosity or frequency of religious attendance; their migration background (i.e. were you born in the United States; yes or no); their self-rated health (poor or fair health v. good, very good or excellent); whether mother and father were living with both of their respective parents during adolescence (yes or no); whether the mother had thought about having an abortion or the father had suggested the mother to have one (yes or no); whether pregnancy affected the relationship between mother and father (worse, same or better); whether the father's last name would be in the birth certificate of the child; and finally whether the mother had worked in the year before the child's birth. Additionally, I included an indicator of the mother's cognitive abilities measured by the Peabody Picture and Vocabulary Test. All these variables were observed at the child's birth, except for the mothers' language ability score which was captured on the 3rd wave. This selection of confounder covariates follows

previous research (Cohen and Pepin 2018; Härkönen 2014; Lichter, Price, and Swigert 2019).

Besides those factors, time-varying confounder covariates may also contribute to explain the departure of the biological father dynamically. This selection of variables is in line with the work of Lee and McLanahan (2015) and others, but I added trajectories for characteristics such as alcohol problems in the family, household composition, public financial help, monetary help from relatives, household wealth, mothers' and fathers' occupational attainment, residential instability, and neighborhood violence:

**Parents' health:** The overall health of the mother was assessed employing a subjective rating. Whether the mother met the criteria for depression and parents' alcohol problems was also included as binary indicators.

**The relation between parents:** Low relationship quality between father and mother based on self-rated relationship quality indicator. I included an indicator of whether the mother reported that her partner had verbally or physically abused her. A binary indicator of whether the biological father had ever been in jail was also included, given that it may affect the relationship between mother and father, although not necessarily break it.

**Living arrangements:** I compared children in nuclear households (where the child lives only with the two biological parents, with or without siblings) v. extended (where the child lives with further extended kin, like grandparents, uncles, aunts, or cousins; with or without siblings) or composite households (where the child lives with people who are unrelated to her; with or without siblings). The presence of siblings from the current or previous relationships, through a binary indicator for the presence of mother-side siblings in the household, as well as a binary indicator for multipartner fertility (whether the mother had children with other men different than the father of the child).

**Socioeconomic indicators:** An indicator of public financial assistance in the form of Temporary Assistance for Needy Families (TANF) or food stamps (yes or no); monetary help from relatives (yes or no); and housing wealth which was constructed based on whether households owned the house they lived, the estimated value of the house minus the debt on the house. For households in which none of its members owned the house where they lived in, a value of zero housing wealth was given; and given that housing wealth was not asked in wave 1, I used data from wave 2 for households that did not change residence between birth and the first year of life). Finally, I included the income to poverty ratio categories (with categories of more than 300%, btw. 200-299%, btw. 100-199%, btw. 50-99%, and btw. 0-49% of the national poverty line) and the equivalent household income at each wave (both of these variables constructed by the survey organizers, which included income from all sources). Parents' occupational attainment was also included. These variables were classified into seven categories: white-collar, high skill (e.g., professional, technical, admin., and executives); services, high skill (e.g., sales, admin. support and services); manual blue-collar (e.g., repair, inspection, and transportation); other low skill (e.g., cleaning, farming and other); self-employed; unemployed; and out of the labor force (OLF).

**Residential instability:** I included changes in residence concerning the previous wave (yes or no), and whether the house the family currently lives in is rented or owned. Finally, to partly capture neighborhood effects, a subjective rating of the neighborhood's safety (very unsafe, unsafe, safe, and very safe) was included. Indicators of whether the mother would be afraid to let her child go outside due to street violence were included in waves for which the neighborhood safety question was not asked.

Previous research has neglected some of these time-varying variables, even though there is evidence suggesting how they may affect family instability

(e.g., Bourdieu 1993; Edin and Kissane 2010; Straus 2017). Therefore, this paper considerably enlarges the set of potential dynamic confounders from what in previous research might correspond to unobserved confounders.

Regarding child wellbeing measures studied in this paper, I looked at a broad selection of child wellbeing indicators, taking advantage of the rich information in the FFCWS:

**Health:** When children were one year old, mothers were asked at each wave whether a doctor had diagnosed the child with asthma. I constructed a dichotomous indicator when mothers responded affirmatively to this question, with no diagnosis of asthma as the reference category. The FFCWS calculated children's BMI, standardized by age and gender, when children were three, five, nine, and fifteen years old. Based on this calculation, I constructed a dichotomous indicator for overweight or obesity in childhood and adolescence, defined as those children's BMIs that were above the 85th percentile of the weight distribution at each year (Cote et al. 2013). Finally, at the 6th wave, to assess healthy behavior, I constructed a dichotomous indicator that captures whether children, already in their teenage years, had tried alcohol, tobacco, or any other drug or substances, with no use of any substance as the reference category.

**Behavioral Ratings and the "non-cognitive" domain:** The socio-emotional domain considers outcomes that relate to problematic behavior in children and adolescents. At the second wave, when children were only one-year-old, the child's emotionality and shyness were assessed employing mothers' ratings on a subset of questions taken from the EAS Temperament Survey for Children (Mathiesen and Tambs 1999). Emotionality refers to irritability or anger, whereas shyness is related to behavior with strangers. Both constructs are associated with later anxiety and depression in young adulthood. I use maternal ratings on six items

(three for emotionality and three for shyness) on a five-level scale to show how characteristic a specific behavior was in her child (from 1 being “not characteristic or typical of your child” until 5 being “very characteristic or typical of your child”). Child Behavior Problems were assessed at the third, fifth, ninth, and fifteenth follow-up surveys. These are based on different subsets of questions from the Child Behavior Checklist 2-3 (Koot et al. 1997). Again, maternal ratings were used on a three-level scale indicating whether a given item was not true (0); sometimes or somewhat true (1); or very true or often true (2) of her child. Two additional measures of behavior in school were included for the time children were adolescents (fifteen years old): a binary indicator for whether the child had ever been expelled or suspended from school; and the trouble at school scale, which consists of a series of statements about situations that may occur to the child in the school context (e.g., paying attention at school, getting along with classmates and teachers, and getting homework done) evaluated by children themselves on a scale from never (=0) to every day (=4). For all constructs based on rating scales, I calculated a total score by adding the individual items in each scale, following previous research and the FFCWS recommendations.

**Cognitive development and educational achievement:** The Peabody Picture and Vocabulary Test (PPVT) was assessed when children were three, five, and nine years old (Hodapp and Gerken 1999). This standardized assessment measures the verbal abilities of children in English and is additionally considered as an indicator of cognitive development. In addition to that, at the sixth wave, teens were asked about the grades they obtained at the most recent grading period in the subjects of science, history, mathematics, and English or language arts (i.e. A, B, C, D or lower, no grade or pass/fail only). Based on these grades I constructed the GPA at age 15. Finally, the event of having ever failed a class, also as a dichotomous

indicator, was considered as an additional measure of educational achievement.

### **Analytical strategy**

First, estimates of the CATE employing BART serve as a contrast to the more conventional approach of ordinary least squares (OLS) baseline adjusted estimates. In these two models, I adjust for all time-invariant confounder covariates, as well as the pre-family transition time-varying covariates observed at  $t - 1$ . I estimate the conditional expectation for each outcome given the treatment and adjusting for baseline confounder covariates, as in  $\mathbb{E}(Y_\tau | Z_\tau, \mathbf{X}_0)$  at each time point  $\tau$ , adjusting for all variables listed in Table 1. For binary indicators of wellbeing, I do this employing linear regression probability models, and for the continuous ones, after standardizing these scales, also linear regression models.

Second, I compare two BART estimates of the CATE to evaluate the relevance of including the confounders' trajectories. This is a comparison of two different CATE estimates employing BART, one adjusted for baseline confounders only, and another one adjusted by all confounders and their trajectories. For each wellbeing outcome  $Y_\tau$  I obtained BART estimates given the treatment  $Z_\tau$  and adjusting for time-invariant at baseline and the history of time-varying covariates  $\bar{\mathbf{X}}_{t-\tau} = (\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_{t-\tau})$ , all measured before the departure of the biological father, which does not include any mediating factor between the departure of the father and the measurement of children's wellbeing. Based on the work of J. Hill and Su (2013), I excluded units whose predicted counterfactual outcome was one standard deviation higher than the observed values for the actual treated units.

Third, given that the effects of family instability may also appear later, I estimate the CATE of earlier father absence on later wellbeing outcomes

using BART, adjusting for the trajectories of confounders. I do this for transitions taking place before 5 years old (early childhood), before 9 years old (middle childhood), and before 15 years old (adolescence). I do this for multiple outcomes studied right after the experience of the family instability, and for later outcomes under an intention-to-treat principle, i.e., regardless of what other family structure changes children might have experienced (Gupta 2011). Although these estimates may underestimate the hypothesized positive effects of father presence on the later outcomes because stable families may not comply with their status and later divorce or separate, they do not, however, affect the estimation of the immediate effects; whereas it would leave unaffected the negative effect of father absence given that the subsequent episodes of family instability of single-mother families should be considered as mediators, and therefore should not be adjusted for in the estimation of the total effect of father absence. Therefore, estimates of the timing effects of father absence presented in this paper should be interpreted as a conservative upper bound for the effect of father absence. But this distinction makes little difference, given the broader statistical tendencies shown by the data.

The average percentage of missing information in the analytical variables was around 18%, which does not deviate much from other longitudinal designs (Huque et al. 2018). To address bias caused by missing information, I use multiple imputation by chained equations with a total of  $M = 20$  imputations, assuming missingness at random (van Buuren and Groothuis-Oudshoorn 2011). The imputation model uses the CART algorithm to find the best set of predictors among the analytical variables to impute the data using all information available across waves, with 10 iterations per imputation (Burgette and Reiter 2010). Only values observed in previous waves are used to impute the missing value at each wave (i.e., observations at time  $t - 1$  are used to impute values at  $t$ ). Linear regression

results make use of the sampling weights using the method of replicate weights, as explained in FFCWS (2019) which accounts for the over-sampling of births to non-married mothers. So far, no method or general recommendations exist for the use of replicate weights with the algorithm BART (Austin, Jembere, and Chiu 2018; DuGoff, Schuler, and Stuart 2014; and Ridgeway et al. 2015). However, this might not pose substantial problems given the variance reducing properties of tree-based methods, such as BART, which may compensate for the remaining bias of not accounting for survey design providing more precise estimates. This form of bias should be expected to be small given the strong reliability of the method to arrive at a fair comparison. The package ‘bartCause’ v. 1.0-4 in R v. 3.6.1 was used to estimate the CATE (J. Hill 2011).

## Results

Figure 1 show the sizes of analytical samples at each time point. The figure shows the reduction in sample sizes as a result of selecting on the family status, which is needed to estimate the effect of interest (i.e., only remaining stable families are considered for the estimation of the later effects).

**Supplementary Materials - Chapter 2** show descriptive statistics for the main baseline characteristics, time-fixed and time-varying variables, broken down by family instability (stable v. unstable) at each of the follow-up time points. The departure of the biological father is more frequent among younger parents who are Black, non-Hispanic; with less than High school; etc., as found by previous research (Härkönen, Bernardi, and Boertien 2017). As children age, differences between families who experience instability on the baseline confounder covariates increases, and children who experience father absence differ on many characteristics from children not exposed to this family transition.

To achieve a fair comparison, it is therefore critical to consider differences in

life courses. Figures 2, 3, and 4 show estimates of the CATE for each child wellbeing dimension, employing BART Baseline (adjusting for the same set of confounders as the linear model, measured at baseline) and BART All (adjusting also for their trajectory before the separation or divorce), with their respective posterior 95% credible regions, organized from highest effect size to lowest; and right above these are the estimates from a linear model adjusted by baseline confounder, and their respective 95% confidence intervals. The first noticeable feature is that BART estimates are rather similar to each other and, for most outcomes, similar to the OLS baseline adjusted models. Therefore, for most outcomes, the linear model would seem to be sufficient to estimate the effects. Adjusting for the trajectory of confounders makes little difference for most outcomes examined in early and middle childhood, with a different pattern for outcomes in adolescence.

Second, the comparison illustrates that BART All’s estimates are smaller and closer to zero than the adjusted estimates illustrated by the OLS baseline adjusted models. The center of the confidence intervals is shifted towards zero for almost all effects. Although confidence intervals overlap to some extent, they are marginally narrower than those of the OLS baseline adjusted models, and the overlap should not be taken to mean that effects are equal. A sizable difference in these estimates can be seen in the shift towards zero. For example, for the outcomes “Asthma at 15 y/o”, “Asthma at 5 y/o”, “Behavioral problems at 15 y/o”, “Behavioral problems at 9 y/o”, “Trouble At School at 15 y/o”, “Shyness 1 y/o”, “Ever suspended from school by 15 y/o”, “GPA at 15 y/o”, and “Ever Failed a class by 15 y/o”, the credible intervals of BART All and BART Baseline contain zero, whereas the OLS baseline adjusted model does not. Therefore, the use of BART leads us to draw different conclusions: If we restrict our analyses to the standard methods, we would have to conclude that father absence has a negative and statistically significant effect on some of the child wellbeing

dimensions, whereas our BART estimates show this is not the case. This finding might have resulted from BART’s strategic search for and selection of confounder covariates, and especially from its nonparametric nature. BART makes results less dependent on researchers’ discretionary choices and on the strong parametric assumptions embedded in standard statistical models, which extrapolate to areas that lack common support.

The third observation that follows, however, is that for outcomes measured at later ages the BART All and the BART Baseline estimates differ on a small magnitude, as seen for the nine and fifteen-year-old outcomes (e.g., “Cognitive Skills at 9 y/o”, “GPA at 15 y/o”, “Ever Failed a class by 15 y/o”, “Behavioral problems at 15 y/o”, and “Trouble At School at 15 y/o”). BART All estimates show a further small shift of the interval towards zero without larger standard errors. Though this difference would not be statistically significant, it is relevant given the overall small effect sizes. This shift may signal that the older the child the more important the trajectories of confounders become because the events that compose the trajectories of confounder covariates are directly experienced by the child. Although the shift is small and there is an important overlap of the confidence intervals, if confounders’ trajectories would not matter, we should not see a shift in the point estimates. Moreover, for “GPA at 15 y/o” the shift of the interval when adjusting for confounders’ trajectories includes zero. This only happens for one of the outcomes here examined, but if we had ignored an adjustment for confounders’ trajectories, we would have reached a different conclusion, namely that a negative exists on this outcome instead of the lack of evidence for it.

Given that the largest effect sizes on children’s wellbeing correspond to the effects of father absence when it occurs close to or during adolescence, the periods of infancy, early, and middle childhood seem unaffected by the departure of the biological father, as shown in other research (Hadfield et al.

2018; Lee and McLanahan 2015), and in contrast to Cavanagh and Huston (2008)’s findings. Given the overall small effect sizes, differences in point estimates are small. However, most wellbeing dimensions examined during adolescence show also no sizeable negative effect, with the sole exception being the higher rate in substance use among fifteen-year-olds, which seems unaffected by the adjustments of BART. This suggests there might be a causal negative effect of the departure of the biological father, on the risk of prior substance use among teenagers who experienced their biological father’s departure from the family unit, even after adjusting for life course selection effects.

To explore matters further, I also estimate the delayed effects of earlier father absence on later wellbeing outcomes using only BART All, as shown in Figure 5 for early childhood, in Figure 6 for middle childhood, and in Figure 7 for adolescence. These figures show point estimates of earlier experiences of father absence on later outcomes, and their 95% posterior credible regions. There is little evidence of delayed effects of father absence during early childhood for transitions occurring before children were five years old, and for transitions taking place before children were 9 years old as well. We see this in the flat curves for the prevalence of asthma and obesity, as well as on the curves of behavioral problems and cognitive skills scores in these age groups, which suggest early father absence effects appear neither in early or in middle childhood.

However, Figure 7 shows a more complex pattern for wellbeing in adolescence. In the health domain, curves for the prevalence of asthma and obesity are flat, with some slight increases when the father’s departure took place during infancy, similar to results for UK children (Goisis, Özcan, and Van Kerm 2019). For substance use by fifteen-year-olds, for which the largest effect size was found, there is evidence of fade-out effects, meaning that earlier experiences of father absence have smaller effects on this

behavior than the more recent transitions; though the posterior credible regions contain the value of zero. For the behavioral domain, effects on Behavioral problems, on trouble at school, and on having ever been suspended from school, results suggest that the departure of the biological father during infancy, early or middle childhood may have effects that appear later on when children reach adolescence, although these differences cannot be distinguished from zero. Among the schooling outcomes, there seems to be an effect of the departure of the biological father during infancy on failing a class. However, this outcome cannot be assigned to a specific developmental stage because it refers to having ever failed a class, though it was measured when children were fifteen years old. For the GPA, results suggest fade-out effects as well. Thus, evidence suggest the effects of father absence may be salient around adolescence, a critical developmental stage, where the confounding adjustment for *life course selection bias* turns the most relevant, in consonance with the more important role of cumulative disadvantages in selecting for divorce or separation.

In summary, evidence suggests that delayed effects of father absence can be to some extent found for behavioral outcomes when children reach adolescence, though not for health-related or scholarly performance, and not during early or middle childhood. Even though most effects are not statistically significant under the null hypothesis of no effects, we can observe the timing effects on the problem behavior dimension of wellbeing for fifteen-year-olds, particularly when transitions took place many years before. This finding provides more confidence for a small and negative causal effect of father absence during this developmental stage.

## Discussion

In contrast to the conclusions of McLanahan, Tach, and Schneider (2013), who summarized the evidence for the causal effects of father absence, my

results suggest that the departure of the biological father does not have substantive negative effects on children’s wellbeing. Exposure to this form of family instability may therefore be considered or interpreted as a marker of life course socioeconomic disadvantage, rather than a cause of negative effects. However, when children in this urban US sample reached adolescence, small negative effects appeared in the behavioral dimension of wellbeing. Although this result requires further investigation, it would correspond to a truly causal effect of father absence on children’s wellbeing. This may imply that small negative effects of father absence are characteristic of adolescence.

However, these findings contradict the family stress theory regarding the effects of father absence, and are not supportive of the family instability hypothesis. If a lack of resources or disruptions in the family system caused by the departure of the biological father brings about negative effects on children’s wellbeing, these should be found for most outcomes across children’s life course, not just for one single outcome observed in adolescence. Although the family instability hypothesis focuses on the joint effect of further episodes of instability, that theory relies on the assumption that the departure of the father has negative effects on children and therefore this paper provides evidence that does not support that hypothesis either. Moreover, estimates on later outcomes are of a similar or smaller magnitude than the estimated effects on outcomes measured right after the experience of father absence, which were all between null to small effects. These small effects are found for families within large cities in the US, for which the FFCWBS was designed. Similar effects could be larger or smaller in other contexts depending on the existence and strength of the safety net supporting families and single mothers (Edin and Kissane 2010). For example, De Vaus et al. (2017) show variability in the effects of divorce on women’s household income across OECD countries, pointing towards a

country’s social security system, social safety nets, and labor market characteristics as important modifiers or moderators of these effects.

Overall, evidence suggests that father absence has no substantive negative effects on early or middle childhood wellbeing indicators. Selection accounts for all the differences in children’s outcomes between those exposed to the departure of the biological father and those children unexposed to this transition. Similarly to studies on fade-out effects of early childhood interventions, the effects of father absence early in childhood on later child outcomes also tend towards zero (Cavanagh and Huston 2008; Bailey et al. 2017). This paper’s results are robust to the inclusion of the observed history of confounder covariates that may affect selection into this form of family instability, which was obtained by adjusting for more confounder covariates. The results are also robust to assumptions about correct model specification, thanks to the use of a non-parametric estimation method designed for causal inference. This leads us to reconsider why changes in family structure would have long-lasting consequences in the lives of children, as to reproduce inequality and disadvantage across generations, when the effects of the departure of the father are small, in most cases not even distinguishable from the null, and especially when effects of targeted developmental interventions instead, with much larger effect sizes, fade out over time (Bailey et al. 2017).

However, despite the more stringent adjustments employed by BART, for the behavioral domain among adolescents I found negative effects of the departure of the biological father. The fade-out plots suggest earlier experiences of father absence when children were 1 and 9 years old appeared to affect problematic behavior when children reached adolescence, as found in other research (Ryan and Claessens 2013; Laird, Nielsen, and Nielsen 2020). Thus, although most of these effects cannot be distinguished from zero, results suggest adolescents exposed to this family transition are more

likely to have tried substances such as alcohol, tobacco, or drugs, to fail a class in school, or to be suspended from school, as well as to show higher scores on the behavioral problem rating scale; net of all previous characteristics that may affect selection into divorce or separation. These outcomes are markers of behavioral or non-cognitive development linked to low self-regulation (Magar, Phillips, and Hosie 2008). Prior research had identified problem behavior as a sensitive domain in which changes in family structure may bring about negative consequences for children (Cavanagh and Fomby 2019; Fomby and Cherlin 2007; and McLanahan, Tach, and Schneider 2013), but the results of the more robust analysis presented in this paper point towards adolescence, and not early childhood, as the developmental stage in which the effects of father absence appear. The consequences of these small differences may turn out to be of substantive importance, given that many crucial educational and work choices take place during adolescence (Dahl et al. 2018; Spengler, Damian, and Roberts 2018). In that regard, research shows that interventions geared towards teenagers' behavior problems may reduce these negative effects (Smithers et al. 2018; Haggerty, McGlynn-Wright, and Klima 2013; and Patton et al. 2018). This would be in line with what certain research calls the building of "character skills" to confront the challenges of growing up into adulthood (Kautz et al. 2014). And as document in the work of McLanahan, Tach, and Schneider (2013), effects on the socio-emotional development in adolescence repeatedly appear in earlier empirical work, and they increase adolescents' risky behaviors which may explain the effects of father absence on high school graduation rates. Notwithstanding, comparing effects across life course stages is difficult because different outcomes might be relevant for different stages, therefore the effects on adolescence may not be comparable to those in infancy. Given that most effect sizes were small and most of them not distinguishable from the null hypothesis of no effect, future research should

explore the extent to which children’s future chances as adults may depend on these small differences, or whether effects fade-out in this case as well.

Similar to the arguments made against the conclusions of Moynihan (1965), policies aimed at promoting stable marriages should foremost address the long list of factors that put families at risk of dissolution, many of which may negatively affect children’s wellbeing as well (Maldonado and Nieuwenhuis 2015). Furthermore, social policies that benefit single mothers in the US have often the unintended effect of discouraging employment and delaying the formation of stable families (Edin and Kissane 2010; Quadagno et al. 1994, 135–54). For example, factors such as affordable housing, employment, and housing stability, and family violence, could not only reduce the risk of divorce or separation but also improve children’s wellbeing without focusing on stabilizing families (Pilkauskas and Micheltore 2019). We could thus consider affordable housing policies as an alternative to promote stable families. Other findings in demographic research are driven to a large extent by selection effects, such as the associations between fertility and family policies in Nordic countries (Andersson 2020), and the marriage wage premium (Ludwig and Brüderl 2018; Killewald and Lundberg 2017). In connection to this, family formation and dissolution are endogenous processes whose determinants remain to a large extent unknown (Manski 1993; Ginther and Pollak 2004, 691–93; Hirschman 2016). These processes are determined by numerous complex interacting factors and therefore are quite unlikely to be affected by narrow policy changes (Cahill 2005).

Despite the advantages of this study, there remain several limitations worth mentioning. First, the sample only captured children born in big US cities and therefore the findings of this paper refer to this urban sample. Second, as most studies based on longitudinal data, attrition may affect estimates in this paper, given unstable families drop out at higher rates than stable families. Attrition, however, does not impact the main conclusion of the

study, which is that better adjustment strategies are needed to make fair comparisons between different family structures because OLS baseline control approaches are too simplistic to adjust for confounding. Although the existence of selection bias may imply the estimated CATEs are closer to the lower bound, all three estimates would be equally affected by attrition, thus not affecting one of these estimates more than another. The design of FFCWS was aware of sample attrition for the study of fragile families, and, therefore, designed an over-sample of births to unmarried couples. Attrition, although still playing a role as in any longitudinal study, would not make up the biggest problem for these findings. Third, another limitation of this study is that the higher leverage on the common causal support assumption enabled by BART, comes at the price of not being able to know what goes on amid interactions between confounder covariates, or which of those confounders is important.

Moreover, as in all observational studies, unobserved variables may continue to play a role. For example, the largest negative effects were found when children reached adolescence, but the 15th year follow-up survey was separated by six years from the previous 9th year follow-up. Therefore, these estimates may continue to be upwardly biased because of the remaining selection bias that I could not adjust for. A conservative interpretation of these results suggests caution, given the overall statistical tendencies found for the other estimates. Future studies, perhaps based on data following children's development on smaller time intervals, could explore whether these found negative effects remain when a better adjustment for trajectories is possible, but such studies often lack many of the confounders used in the adjustment set in this paper. Therefore, we require stronger evidence to support claims of father absence affecting problematic behavior in adolescence. Finally, father absence is never a complete absence, and some form of visiting or custodial arrangements is in

place. Therefore, children’s actual lived experiences of father absence may matter, and depending on custodial arrangements and actual time spent together with fathers, the effects of father absence may be moderated. Such analysis would require a different methodological approach to the one employed in this paper, which is left for further research.

Compared to previous studies on the same effect, the use of BART has allowed me to include many more confounder covariates than previous studies had done, and to test whether effects hold employing nonparametric methods. This study shows how findings of previous research are affected by the strong and untestable parametric assumptions of statistical models that are used to show how the effects of father absence affect children’s well-being, in linear ways (Abbott 1988). This paper overcomes these limitations by adjusting for multiple-way, non-linear, and time-dependent interactions among confounder covariates and their trajectories, as observed before the departure of the biological father. Here I have compared multiple child wellbeing outcomes that were expected to be affected by father absence, as hypothesized and explored in this literature, which allows for a ready comparison of wellbeing domains. To my knowledge, besides systematic reviews of the literature and a study employing BART and the FFCWS in a prediction modeling task (Carnegie and Wu 2019), this is the first study to look at this number of outcomes from a causal inference perspective, and to report multiple estimates. This paper shows that with one single exception, effects in the health, behavioral and cognitive domains tend towards the null hypothesis of no effects. *Life course selection bias* seems to play a small though noticeable role for outcomes examined close to or during adolescence, but not on those in early or middle childhood. This is, however, in line with the role of cumulative disadvantages not only on outcomes but also crucially on selection processes for divorce or separation. This paper contributes to a better and more fair comparison between

children who experienced this type of family transition and those who were not. Future similar studies for further family transitions, such as the effects of step-families on children’s well-being (Saint-Jacques et al. 2017), as well as other topics within family sociology, could explore the *life course selection bias* in other applications. The life course perspective suggests strong interactions between the distant past and the present, which implies that more flexible approaches are needed if accounting for complex selection bias mechanisms is necessary. Given that many demographic behaviors are endogenous (see Ginther and Pollak 2004; Manski 1993), demographic causal inference should consider more flexible approaches such as BART to better inform contemporary debates on its effects, especially those involving longer sequences of causality. As shown in this paper, machine learning plus causal inference methods might provide a convenient solution to achieve this (Dorie et al. 2019). For the case of BART, introductory material to this method exists for researchers interested in exploring other topics for which such complex forms of confounding, as stated in the *life course selection bias*, might make up a form of dynamic confounding (Dorie et al. 2019; J. L. Hill 2011; J. Hill, Linero, and Murray 2020; Tan and Roy 2019).

Precisely because family dynamics imply complex selection processes, as processual sociological accounts of social phenomena invite us to consider (Abbott 2016), more flexible approaches such as BART could advance the life course perspective in family sociology. Even though the departure of the biological father out of the family unit does not capture the effect of a complex concept such as family instability, which is measured as a time-varying exposure, these results highlight how dynamic confounding and causal inference assumptions may compromise even the simplest of effects that are held as supportive of the family instability hypothesis, i.e., the effect of the first family transition. Moreover, if father absence does not have negative effects on children’s well-being, then the mechanisms

explaining this are of great interest because they highlight the potential for family resilience as an adjustment strategy against negative life course events (Seltzer 2019). The null-findings presented in this paper may be the consequence of families and children's resilience and adaptation to cope with changes in their environments (Kelly and Emery 2003; Hetherington and Stanley-Hagan 1999), which is an important topic to be addressed by future research on father absence and the family instability hypothesis.

## Tables and Figures

*Table*

Table 1: Description of variables used in chapter 2

Variables	Min	Max	N	Description	Waves
Child's verbal ability (PPVT)	40	137	964	Peabody Vocabulary Test (PPVT/TVIP) score	3rd, 5th, 9th years
Child's combined GPA	1.00	4.00	1199	It includes the subjects of English, Math, History or Social Studies, and Science	15th year
Ever failed a class	1	2	1422	Ever failed a class in school? (yes/no)	15th year
Emotionality	3	15	1831	Emotionality - EAS Temperament Survey for Children: Parental Ratings - the tendency to become aroused easily and intensely, broad measure of distress in the very young infant, associated to fear and anger tendencies in the older child.	1st year
Shyness	3	15	1827	Shyness - EAS Temperament Survey for Children - a tendency towards inhibition and awkward behavior in the young child	1st year
Child behavior problems	0	80	1156	Child Behavior Problems (CBCL) including both internalizing and externalizing problems	3rd, 5th, 9th, and 15th years
Trouble at school scale	0	8	1411	Trouble at School in four areas getting along with teachers, paying attention in school, getting homework done, and getting along with other students (0 = Never to 4 = Every day)	15th year
Ever suspended from school	1	2	1423	Ever been suspended or expelled in past 2 years? (yes/no)	15th year
Diagnosed with asthma	1	2	1628	Has a health care professional ever told you child has asthma? (yes/no)	3rd, 5th, 9th, and 15th years
Overweight	1	2	981	Child Body Mass Index standardized by age and gender, binary indicator if child's BMI is higher than the 85th percentile (yes/no)	3rd, 5th, 9th, and 15th years
Use of substances	1	2	1452	Ever smoked an entire cigarette? Drank alcohol more than two times without parents? Tried marijuana? Tried other illegal drugs besides marijuana?	15th year
Child's gender	0	1	2055	Gender assigned at birth (boy/girl)	Baseline
Low birth weight	0	1	1999	Was the child diagnosed with low birth weight? (yes/no)	Baseline
Mother drank alcohol during pregnancy	1	5	2049	During the pregnancy, how often did you drink alcohol? (1=Everyday to 5=Never)	Baseline

Table 1: Description of variables used in chapter 2 (*continued*)

Variables	Min	Max	N	Description	Waves
Mother smoked during pregnancy	1	4	2050	During the pregnancy, how many cigarettes did you smoke? (1=more than packages per day - 4=None)	Baseline
Mother took drugs during pregnancy	1	5	2051	During the pregnancy, how often did you use drugs? (1=Everyday to 5=Never)	Baseline
Father's last name on birth certificate	0	1	2042	Will the baby (babies) have the father's last name? (yes/no)	Baseline
Mother's age at child's birth	15	43	2054	Mother's age (years)	Baseline
Father's age at child's birth	16	53	1858	Father's age (years)	Baseline
Mother's education at child's birth	1	4	2053	Mother's education (1=less than highschool, 2=highschool or equivalent, 3=some college, technical education, or 4=college or graduate)	Baseline
Mother's verbal ability (PPVT)	40	139	929	Mother's or primary care giver PPVT - Standardized score	3rd year
Father's education at child's birth	1	4	2046	Father's education (1=less than highschool, 2=highschool or equivalent, 3=some college, technical education, or 4=college or graduate)	Baseline
Mother's race	1	4	2051	Mother's race 1= white, non-hispanic; 2 black, non-hispanic; 3=hispanic; 4=other	Baseline
Father's race	1	4	2054	Father's race 1= white, non-hispanic; 2 black, non-hispanic; 3=hispanic; 4=other	Baseline
Mother is US citizen	0	1	2053	Was the mother born in the U.S.? (yes/no)	Baseline
Father is US citizen	0	1	1856	Was the father born in the U.S.? (yes/no)	Baseline
Mother's religiosity	0	1	2052	How often does mother attend religious services? (at least once a year vs. never)	Baseline
Father's religiosity	0	1	1857	How often does father attend religious services? (at least once a year vs. never)	Baseline
Mother lived with both her parents	0	1	2035	Was the mother living with both of her biological parents at age 15? (yes/no)	Baseline

Table 1: Description of variables used in chapter 2 (*continued*)

Variables	Min	Max	N	Description	Waves
Father lived with both his parents	0	1	1847	Was the father living with both of her biological parents at age 15? (yes/no)	Baseline
Mother thought about abortion	0	1	2051	When mother found out she was pregnant, did she think about having an abortion? or the father suggested her to have an abortion?	Baseline
Father thought about abortion	0	1	1852	When father found out the biological mother was pregnant, did he think about her having an abortion?	Baseline
Mother's overall health	0	1	2054	How is the mother's health? (1=Great to 5=Poor)	Baseline, 3rd, 5th, 9th, and 15th years
Mother-father relationship quality	1	3	2040	After mother found out she was pregnant, how did mother's relationship with biological father change? (better, worse, same) And after baseline, how is mother's relationship with child's father? (1=excellent to 5=Very bad)	Baseline, 3rd, 5th, 9th, and 15th years
Mother with alcohol/drug problems	0	1	2052	In last year, have alcohol/drugs interfered with mother's work/relationships? (yes/no)	Baseline, 3rd, 5th, 9th, and 15th years
Father with alcohol/drug problems	0	1	1855	In last year, have alcohol/drugs interfered with father's work/relationships? (yes/no)	Baseline, 3rd, 5th, 9th, and 15th years
Father has been in jail	0	1	2053	Both mother and father report that father was in jail at each interview (yes/no)	Baseline, 3rd, 5th, 9th, and 15th years
Violence against the mother	0	1	1955	Frequency that father hit or slaps mother when he is angry, insults or criticizes her? (if this ever happened one, else zero)	Baseline, 3rd, 5th, 9th, and 15th years
Welfare/food stamps/TANF recipient	0	1	2039	In last year, did the mother have income from public assistance/welfare/food stamps/TANF? (yes,no)	Baseline, 3rd, 5th, 9th, and 15th years
Financial assistance from other family members	0	1	2050	Have you receive financial support from anyone besides biological father?	Baseline, 3rd, 5th, 9th, and 15th years
Father's socio-occupational category	1	7	1848	What sort of work does/did father do in his current/most recent job? (white collar, high skill; services, high skill; manual blue collar; other low skill; self-employed; unemployed; or out of the labor force)	Baseline, 3rd, 5th, 9th, and 15th years

Table 1: Description of variables used in chapter 2 (*continued*)

Variables	Min	Max	N	Description	Waves
Mother's socio-occupational category	0	1	2055	What sort of work does/did mother do in his current/most recent job? (white collar, high skill; services, high skill; manual blue collar; other low skill; self-employed; unemployed; or out of the labor force)	Baseline, 3rd, 5th, 9th, and 15th years
Neighborhood violence level	1	4	2045	How safe are the streets around your home at night or frequency of gang activity? (1= Very safe to 5=Very unsafe)	Baseline, 3rd, 5th, 9th, and 15th years
Mother has moved from previous house	0	1	1583	Has the mother moved houses since child was born or since last interview? (yes/no)	Baseline, 3rd, 5th, 9th, and 15th years
Mother lives in a rented house/apartment	0	1	2038	Is the home/apartment were mother currently resides owned/rented?	Baseline, 3rd, 5th, 9th, and 15th years
Non-Nuclear family structure	0	1	2055	A synthetic indicator created from household members information (dichotomous indicator if non-nuclear family structure is present)	Baseline, 3rd, 5th, 9th, and 15th years
Child's siblings live in the same household	0	1	2055	A synthetic indicator created from household members information (dichotomous indicator if at least one child's sibling is present)	Baseline, 3rd, 5th, 9th, and 15th years
Housing wealth	-315002	5000000	1041	Net housing wealth (difference between the value the house could be sold minus what is owed to the bank)	Baseline, 3rd, 5th, 9th, and 15th years
Equivalized household income	0.0000	94575.5320	2041	Household income combining all sources of income, divided by the square root of the household size	Baseline, 3rd, 5th, 9th, and 15th years
Poverty categories	1	5	2055	Poverty categories constructed by FFCWB based on mother's household income/poverty threshold ratio (1 = 0-49% to 5 = 300%+)	Baseline, 3rd, 5th, 9th, and 15th years
Mother's depression	0	1	1839	Mother meets depression criteria (liberal) at one-year (based on the CIDI questionnaire)	3rd, 5th, 9th, and 15th years
Child's age (at second wave)	9	30	1837	Baby's age at time of mother's one-year interview	3rd, 5th, 9th, and 15th years
Multipartner fertility indicator	0	1	1833	Mother has children by man other than the biological father of child	3rd, 5th, 9th, and 15th years

*Figures*

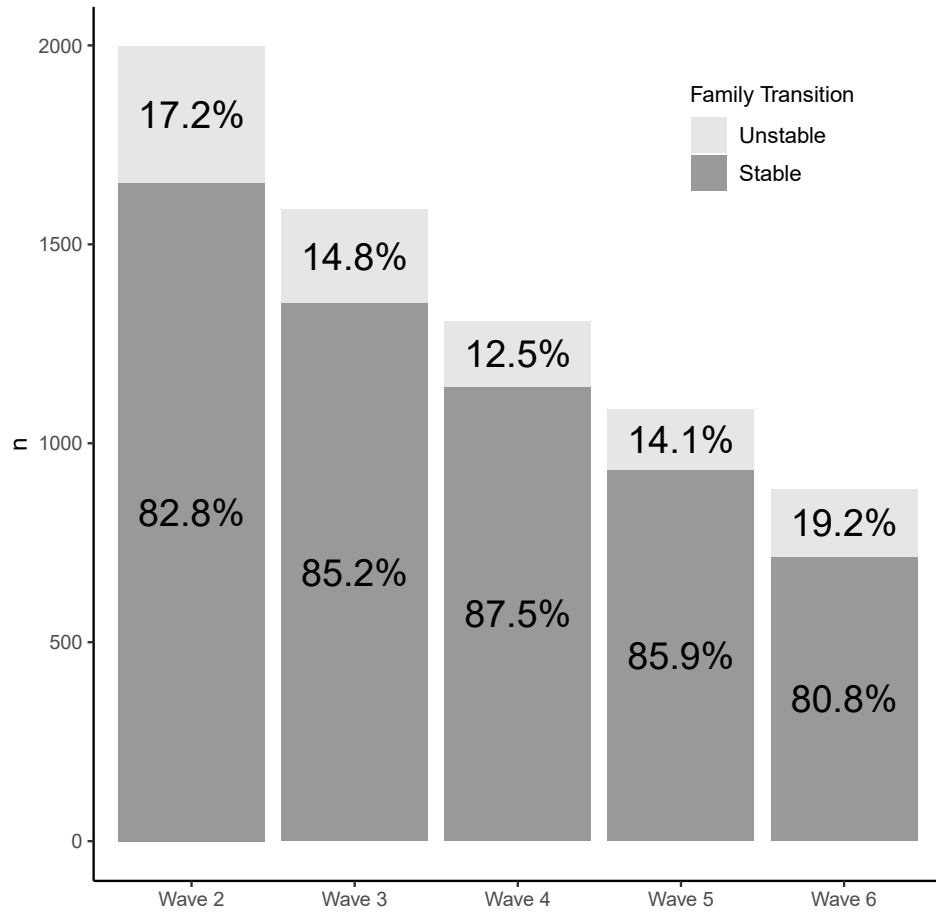


Figure 1: Sample Sizes at Each Wave by Family Instability, FFCWS

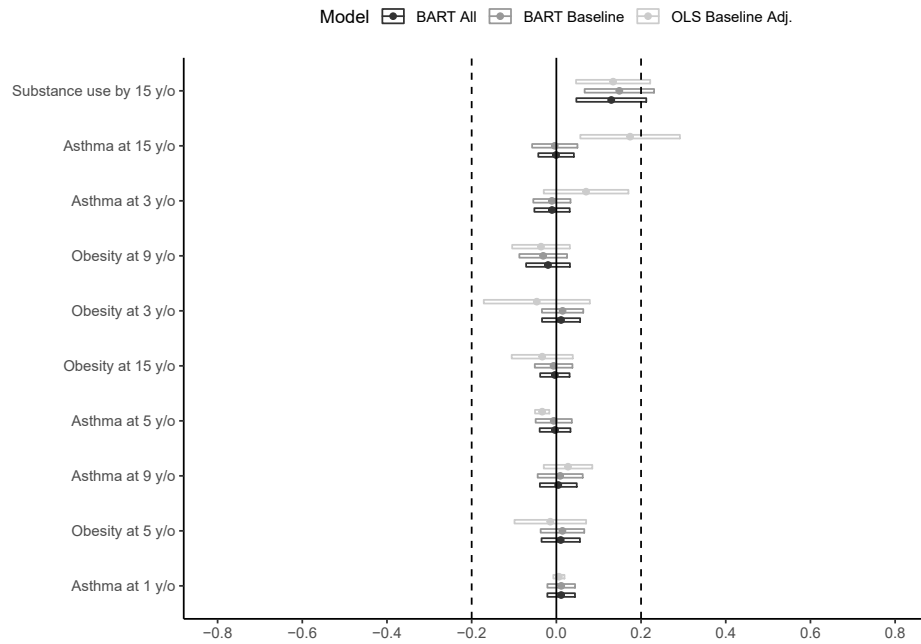


Figure 2: CATE of Departure of Bio. Father on Health wellbeing Outcomes, FFCWS

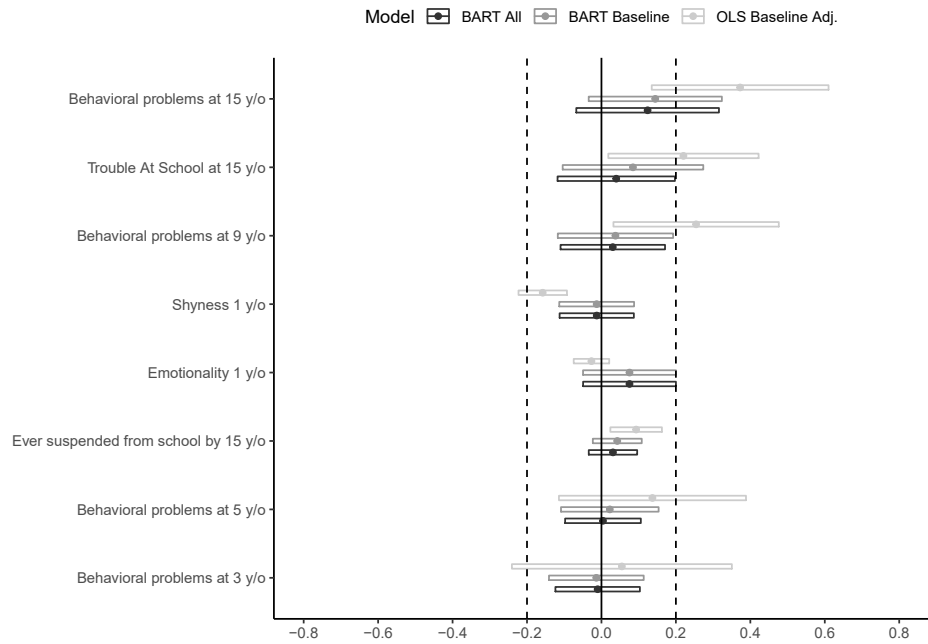


Figure 3: CATE of Departure of Bio. Father on Behavioral/Noncognitive wellbeing Outcomes, FFCWS

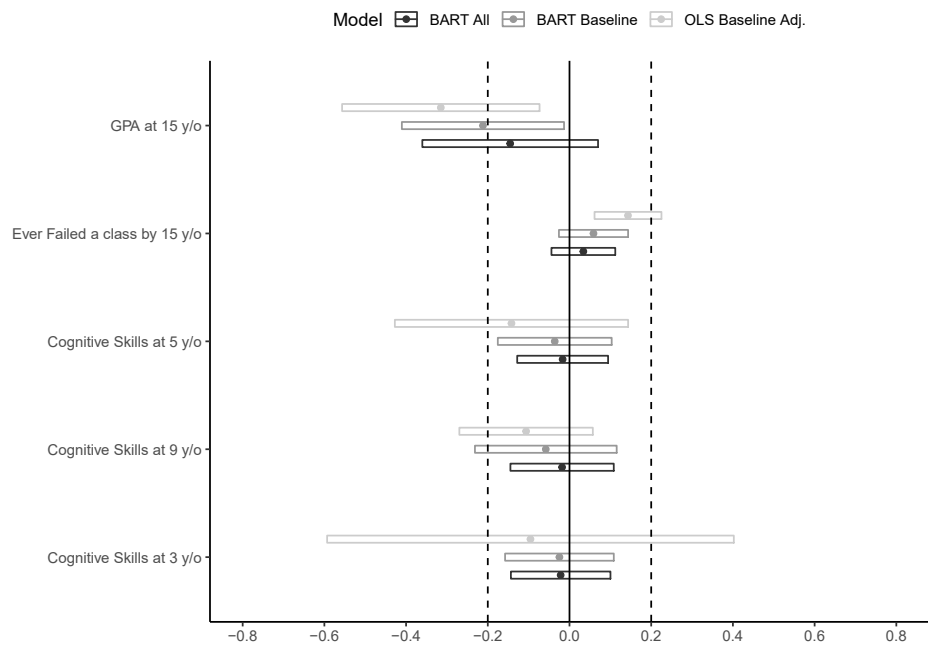


Figure 4: CATE of Departure of Bio. Father on Cognitive/Schooling wellbeing Outcomes, FFCWS

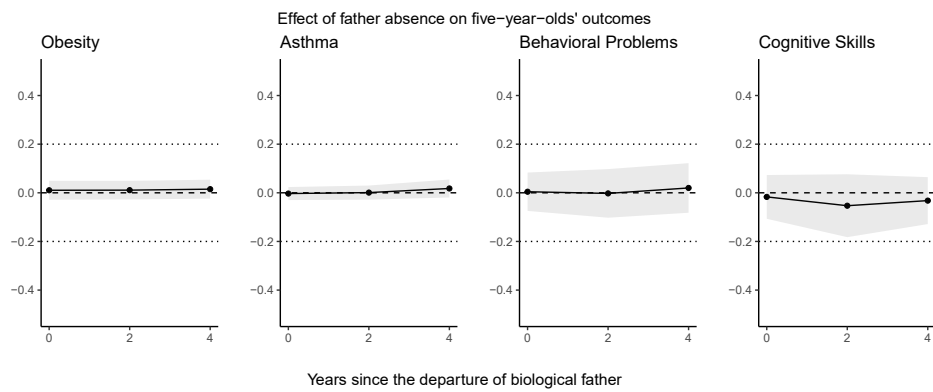


Figure 5: Timing - CATE of Earlier Father Departure on Later Early Childhood wellbeing Outcomes, FFCWS

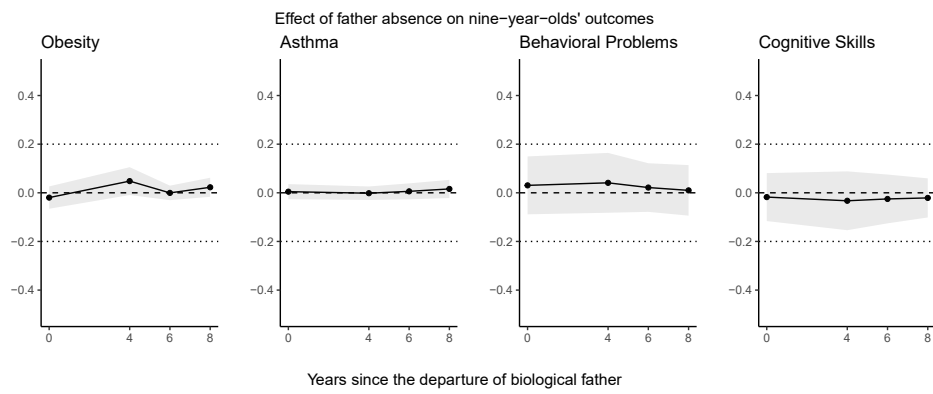


Figure 6: Timing - CATE of Earlier Father Departure on Later Middle Childhood wellbeing Outcomes, FFCWS

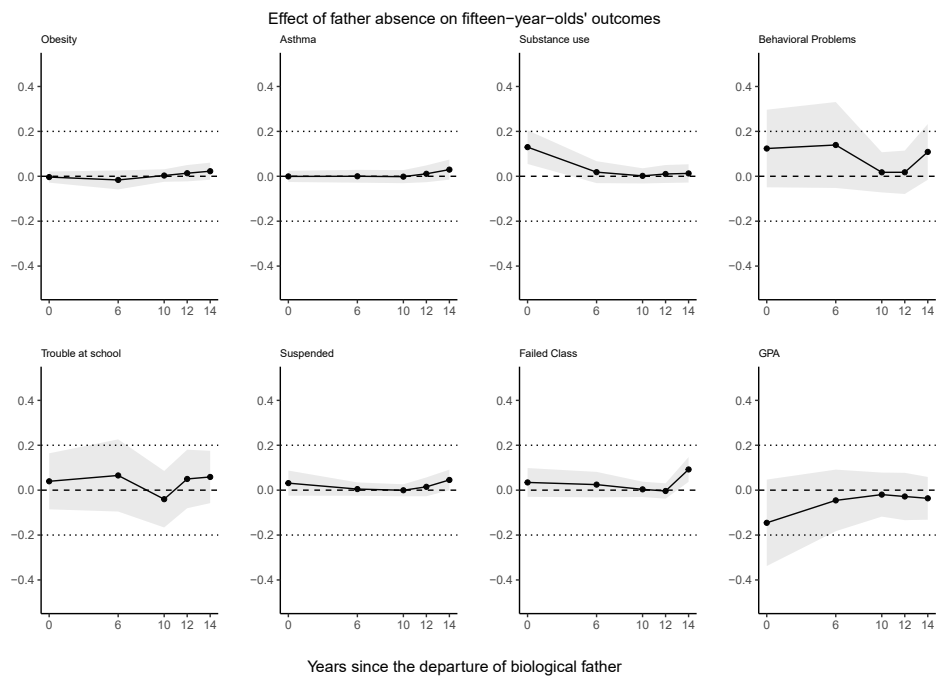


Figure 7: Timing - CATE of Earlier Father Departure on Later Adolescence wellbeing Outcomes, FFCWS

## **CHAPTER 3 - The complex effects of family instability on adolescent problem behavior in a U.S. birth cohort**

Almost three decades ago, Wu and Martinson (1993) advocated for the use of dynamic measures of family structure to examine the family instability or change hypothesis. This hypothesis claims there are negative effects on children's wellbeing associated with experiencing repeated changes in family structure during childhood (Cavanagh and Fomby 2019). The main argument Wu and Martinson (1993) made at the time was that the use of static measures of family structure was inappropriate to capture an essentially dynamic effect (Amato and Keith 1991; Wells and Rankin 1991). Therefore, the approach of comparing the wellbeing of children living in different family structures, adjusting for potential 'control variables,' was insufficient. Instead, empirical studies of family instability, and in general family complexity, required dynamic measures of family experiences. This led researchers to focus on the number of changes in family structure, or cumulative instability (Cavanagh and Huston 2008), which became one of the main measures of family instability used in demographic research – together with the type of family transition, distinguishing relationship formation from its dissolution. Despite more recent findings that the number of family changes is negatively associated with children's wellbeing (Lee and McLanahan 2015; McLanahan, Tach, and Schneider 2013; McLanahan and Percheski 2008), and in particular child behavior problems, research on family dynamics has not fully address the empirical challenges of estimating this dynamic effect.

The approach of Wu and Martinson (1993), and of others who followed (Fomby and Cherlin 2007), though important breakthroughs, were left incomplete. One reason why this is the case is that changes in family

structure are causal factors affecting the probability of further family changes (Crosnoe et al. 2014). For instance, separation or divorce leads to, among other things, residential changes, temporarily moving in with extended kin, reductions in household income, as well as to increases in divorced mothers' labor force participation. These events explain an important chunk of the effect of family instability (i.e., as mediators of family changes, as suggested by Brand et al. 2019). However, those intermediate effects may also affect the probability of further family changes, they affect the risk of finding a new partner (Edin and Kefalas 2005), as may result from moving to a new neighborhood or from entering new jobs with precarious working conditions (Cohen 2002). Multiple studies provide examples of this type of causal chains that likely follow family transitions (Hall, Iceland, and Yi 2019; Harding et al. 2010; Jacoby et al. 2017; Mikolai and Kulu 2018; and K. L. Perkins 2017).

If family changes at different time points affect children's problem behavior – though with differences in their effect sizes depending on the type of family change (Hadfield et al. 2018; Saint-Jacques et al. 2017) – the intermediate events that are caused by a family change mediate the effect of an initial transition, but confound the association between a second family change and children's wellbeing. For this reason, the effect of number of family transitions experienced by children, for example, two transitions compared to no transitions ('intact' families), cannot be estimated by traditional methods – as reviewed, for example, in McLanahan, Tach, and Schneider (2013) or Härkönen, Bernardi, and Boertien (2017) – without incurring in treatment-confounder feedback bias (Hernan and Robbins 2020; Hernán, Brumback, and Robins 2000).

Assessing whether the number of family structure changes impacts children's problem behavior requires addressing dynamic confounding, especially treatment-confounder feedback bias, and, in particular, should place special

care to the timing of the effect. One way of addressing all these issues is through the use of G-methods, such as Marginal structural models (MSM) with inverse probability of treatment weighting (IPTW) (see Hernán and Robins 2006; Mansournia et al. 2017). The use of MSM has been fairly limited in this research area. For example, Lee and McLanahan (2015) employed this method to estimate the effect of number of family transitions on children’s cognitive development and externalizing and internalizing problem behaviors, with findings suggesting that only White boys problem behavior was affected by family instability net of dynamic confounding. Similarly, K. L. Perkins (2019) employed MSM to estimate the effects of household composition change – including though not limited to changes in the relationship status of parents – on educational attainment, finding here negative effects too. Finally, Harvey (2020) studied how doubling up – residing with extended kin or non-kin adults – shaped children’s life chances, as reflected in lower young-adult health and lower educational attainment impacts, equally employing MSM. The findings of these previous studies are suggestive of negative effects of various forms of instability (i.e., family structure, household composition, and doubling-up).

However, one shortcoming of this research, and another reason why conventional methods, is that they did not consider jointly the timing of these changes (TenEyck, Knox, and El Sayed 2021; Cavanagh and Huston 2008). The timing of the effects matters too, and most studies did not address the effect of more than three changes in family structure, which are empirically observed. Family transitions can occur from birth till the child 18th birthday, but, at least hypothetically, it makes a difference when those changes occur. One, two, three or more changes can take place early in the child’s life, right before adolescence, or can be ‘equally spaced’ throughout childhood. The same number of family changes sustained by two different children can therefore refer to fundamentally different experiences from the

child perspective. For example, Womack et al. (2022) show that both increasing and declining levels of instability are not associated with high externalizing behaviors in adolescence, but are so in childhood – their methodological approach, however, does not differentiate the timing of these effects (Cavanagh and Huston 2008).

Given that MSM do not have a single unique counterfactual trajectory that one can compare stable families to, it is up to the researcher to select which comparison to make. A major shortcoming of previous studies is that it is unclear what is being compared to what: most studies compare different degrees of instability to the most stable or normative ones (i.e., family trajectories without family changes). Although important, this is not the only comparison, nor necessarily the most interesting one we can make. More importantly, for example, Lee and McLanahan (2015) adjusted for a highly restrictive set of time-varying mediators/confounders affected by family change and affecting future family change, which was an important advancement. However, these authors left open other biasing paths by not including other observed time-dependent confounders, such as housing.

In this paper, I present a similar estimation of the effect of changes in childhood family structure using a doubly robust estimation of MSM employing IPTW and considering a larger set of time-dependent confounders. The life course cumulative and pathway models, as explained in Mishra et al. (2009), are tested against the data. Similarly to other causal inference methods, MSM estimate the effects of counterfactual or hypothetical interventions (De Stavola, Herle, and Pickles 2022), and their parameters have a causal interpretation based on a set of assumptions. One crucial assumption behind MSM based on IPTW is that models for the exposure at each time point ought to be correctly specified. In contrast to previous research, I estimate these probabilities employing Bayesian Additive Regression Trees (BART), a machine learning algorithm that is

more flexible and better suited to predictive modeling and causal inference (see, for an overview, Tan and Roy 2019; J. L. Hill 2011), and which has the advantage of considering all potential non-linear interactions between time-dependent confounders.

## **Background**

### **Problem behavior in a cohort of American children**

One dimension of children’s wellbeing on which the effects of family instability have been consistently found is socioemotional development (Cavanagh and Huston 2008; Lee and McLanahan 2015), which relates closely to problem behavior. In contrast to cognitive development, where evidence for effects is scant, previous studies have found that children who experienced repeated changes in their family structure had higher reports of problem behavior than children growing up in stable families (Cavanagh and Huston 2006).

Problem behavior, or socioemotional skills, is a child wellbeing dimension related to the overall cognitive functioning of the child. This dimension taps into various cognitive fundamental skills in child development, such as self-control, the management of emotions, and planning, etc. (Bongers et al. 2003). The problem behavior checklist (Koot et al. 1997) – an indicator used to approach this dimension – roughly captures self-regulatory skills in the child at different ages, by measuring behavioral internalizing and externalizing traits associated with it (i.e., through maternal, teacher, or independent observer ratings). For example, internalizing problems is captured through multiple indicators of withdrawn behavior or behaviors related to anxiety and depression, such as worrying too much, feelings of guilt, crying a lot, or feeling worthless. Externalizing problems, in turn, refer to behavior that can be considered as aggressive, such as arguing a lot,

being disobedient at home or school, destroying things, or physically harming others, and also rule breaking behavior more generally.

In addition to family instability (Wallenborn et al. 2019; Lee and McLanahan 2015), problem behavior in children has been shown to be influenced by neighborhood disorder (Pei et al. 2019), mother's ethnic-racial identity (Lazarevic, Toledo, and Wiggins 2020), food insecurity (Hobbs and King 2018), and other socioeconomic determinants related to poverty. Such contextual factors are further potential sources of endogeneity (i.e. selection bias) that confound the association between number of changes in family structure and children's problem behavior. These factors further affect the probability of selecting into multiple changes in family structure, but are equally affected by changes in family structure too.

### **Family instability: the accumulation and timing of its effects**

A key dimension of childhood family structure trajectories are the number of changes or transitions (Cavanagh and Huston 2008), which some research suggest are far more substantial for children's wellbeing than the type of family transition experienced (Fomby and Cherlin 2007; Lee and McLanahan 2015). Studies suggest that the departure of the biological father has more detrimental effects than the entrance of a social father, but both transitions are negatively associated with children's wellbeing.

Although growing up in stable family arrangements has been shown to be beneficial in terms of children's socioemotional wellbeing, empirical evidence for the effects of these two types of family transition is mixed.

The balance of recent literature reviews seems to support the claim of small negative effects (Amato 2010; F. Bernardi et al. 2013; Esping-Andersen 2016; Hadfield et al. 2018; Härkönen, Bernardi, and Boertien 2017; McLanahan, Tach, and Schneider 2013; McWayne et al. 2013; Saint-Jacques

et al. 2017). Changes in family structure bring about stressful and negative events in the life-course of adults and their children. Presumably more so for children with a racial-ethnic background and from more recent cohorts, among which family instability is more prevalent (Census Bureau 2019; Smock and Schwartz 2020), and particularly so for U.S. children (Fomby and Johnson 2022; Cherlin 2010). In theory, changes in family structure negatively affect children because their parents cannot fully benefit from the gains of marriage (G. S. Becker 1973; Schultz 1974; Browning, Chiappori, and Weiss 2014). Therefore, family instability would be an important contributing factor in the perpetuation of poverty and inequality in the U.S. (McLanahan 2004).

From a dynamic perspective, the family instability hypothesis focuses on the reinforcing processes that, following a change in family structure, may negatively affect children's wellbeing. For example, single-parent families have, on average, less financial and time resources than families in which the two biological parents are present; whereas blended, complex or stepparent families sub-optimally invest in children as a result of the stepfathers' lack of incentives to care for stepchildren's human capital formation (Browning, Chiappori, and Weiss 2014) [pp. 438-470]. Furthermore, the quality of inter-parental and parent-child relationships is also affected by family instability. Disruptions in the roles and routines of parents and their children caused by divorce or repartnering require adaptation and adjustment from the complete family system (Masarik and Conger 2017), therefore becoming an additional source of stress. Moreover, in families resulting from changes in family structure, the roles of the single-parent or the stepparent ought to be negotiated anew, which creates the potential for conflict between parents and children, who, in addition, presumably lack an extra paternal source of authority or attachment (Sigle-Rushton and McLanahan 2004). Therefore parents' relationship quality, as well as their

mental health, and parenting quality are presumed to be lower among families experiencing instability (Hadfield et al. 2018). For these reasons, most previous research has suggested that changes in family structure, and, in particular, the frequency or number of changes, are *causes* of negative effects in the life-course of children.

However, the life course theory suggest the influence of changes in family structure likely differs by the child's age at the time the family transition takes place (Elder Jr, Shanahan, and Jennings 2015; Cavanagh and Huston 2008). Each additional family transition is associated with cumulative stress and negative effects on the child (Gosselin, Babchishin, and Romano 2015). But some research suggests that the different needs and levels of dependency of children vis-à-vis adolescents could explain specific timing effects (Hadfield et al. 2018, 101–3). Despite the difficulty of capturing the timing aspect with survey data of high temporal granularity, some studies suggest these effects are larger in early childhood than in adolescence (Ryan and Claessens 2013; Fomby 2013), whereas other studies on specific transitions find only small negative effects for transitions experienced during or close to adolescence (Rodriguez S. 2021).

### **Dynamic selection into family instability from a life-course perspective**

As highlighted by Wu and Martinson (1993), family instability should be conceived of as a time-varying exposure unfolding over children's life-course. In this regard, family dynamics are the result of heterogeneous social and economic circumstances affecting family formation/dissolution processes (Seltzer 2019). Especially in the US, family dynamics exhibit a high level of complexity (Fomby and Johnson 2022; Brown, Manning, and Stykes 2015; VanOrman and Scommegna 2016). The life-course perspective suggest this complexity results from counteracting processes that also take place after

experiencing changes in family structure (L. Bernardi, Huinink, and Settersten Jr 2019, 3–7).

These processes may have important consequences for family theories that often assume effects of family instability to follow the life course cumulative instability model. However, zooming in on the causal pathways leading from family instability to worse outcomes in children reveals substantial nuances. Family instability activates what could be conceived as interlocking trajectories in various other domains of the life-course of parents and their children that affect each other in dynamic ways: both in negative and therefore reinforcing ways; as well as positive, and therefore counteracting ways. Critics of the family instability hypothesis argue that the negative association between family instability and children's wellbeing is the result of parents' characteristics and events taking place in their life course, making some children more prone to experience changes in the relationship status of their parents than other children (Härkönen, Bernardi, and Boertien 2017; Rodriguez S. 2021). The dynamic selection hypothesis proposes that socioeconomic factors such as low income, unemployment, or low education – which are associated with changes in family structure – drive the often empirically found negative association, in what causal inference refers to as dynamic confounding (Jackson 2016). By adjusting for selection into specific family structures (F. Bernardi and Boertien 2017), effects are substantially reduced and are in general not statistically significant anymore. Other research argues that, although confounding does exist in this association, after appropriately adjusting for it, there remains a negative, albeit smaller effect of family instability on children's wellbeing (McLanahan, Tach, and Schneider 2013). Moreover, the evidence seems consistent across various children's wellbeing dimensions and other Western developed countries show similar effects (Amato 2010), which lends more credence to the family instability hypothesis.

A further research strand within this literature has focused on the events that trigger those changes in family structure, as well as on the effects of those family structure changes on further family instability (Cooper and Pugh 2020). For example, in response to changes in family structure, single mothers make decisions to accommodate to new circumstances and respond to the changes in family resources (K. L. Perkins 2017; Mikolai and Kulu 2018; Edin and Kefalas 2005). Divorced mothers may strategically reconfigure the composition of their families by moving in with extended kin; by changing their residence (Jæger 2012; Raley et al. 2019; Desmond and Perkins 2016); by cutting on unnecessary expenses and by re-organizing their schedules; or even by finding a new partner. Mothers who experience a disruption in their partnership trajectories – and fathers as well – may find ways of compensating for the lost family resources after a change in family structure took place (Erola and Kilpi-Jakonen 2017, 9–14); but these changes are likely to further affect the trajectories of family formation leading to more or less future family instability, thus making previous family instability a causal factor of future family instability (Cohen and Pepin 2018; Lyngstad 2011). Such feedback mechanisms are characteristic of complex family dynamics, and, importantly, they also shed light on the problems of dynamic confounding for the estimation of effects of a time-varying demographic behavior such as repeated changes in family structure (Hernan and Robbins 2020, 247–55).

## **Methods and Data**

### **Methods**

In other words, changes in family structure make further changes more likely because of two reasons. First, in most cases, repartnering can only occur after a separation or divorce has first taken place. Second, because the mechanisms triggered by the first change in family structure affects the

chances of further family changes. Family research early recognized this feature of family behavior as endogeneity (Ginther and Pollak 2004), which implies – for estimates of causal effects of family instability on children’s wellbeing – that there is dynamic confounding, given that not all families are equally likely to experience subsequent family changes. In general, whether a child will experience family instability in their life-course is the result of a dynamic treatment trajectory. The directed acyclic graph (DAG) in Figure 8 portrays the interactions implied by such dynamic trajectories, and where dynamic confounding is observed [Pearl (2009); p.65-105]. The DAG shows a measure of children’s problem behavior  $Y_{15}$  – when the child was already 15 years old – as affected by the trajectory of family instability experienced by the child ( $\bar{F}_{15} = \{F_0, F_1, F_3, F_5, F_9, F_{15}\}$ ) and multiple confounding factors. These factors affect family change at all time points, and encompass baseline time-invariant characteristics ( $W_0$ , such as parental education, race-ethnicity, age of parents at child’s birth, etc.); and also the trajectory of time-varying characteristics ( $\bar{V}_{15} = \{V_0, V_1, V_3, V_5, V_9, V_{15}\}$ , such as household income, residence, employment, etc.); unobserved variables  $U^Y$ , such as social networks – that in turn, for simplification, may affect all processes in the sequence. Therefore, problem behavior is in this DAG a function  $g$  (unknown) of multiple events in the child’s life course  $Y_{15} = g(W_0, \bar{V}_{15}, \bar{F}_{15}, U^Y)$ .

Following the data structure of the FFCWS, the DAG represents the main theoretical ideas behind the family instability and the selection hypotheses. The family instability hypothesis focuses on the changes in resources captured by  $\bar{V}_{15}$  (Brown, Manning, and Stykes 2015; VanOrman and Scommegna 2016), as well as the mediating effects of parenting related variables (Masarik and Conger 2017), that are affected by family changes. These parenting mediators, as well as other mediating mechanisms not affecting family changes, are excluded from the DAG given that the focus is

on estimating the joint effect of repeated family changes. The selection hypothesis makes emphasis, in turn, on the arrows that go from the unobserved factors  $U^Y$  and the observed baseline characteristics  $W_0$  and  $V_0$ , to the exposure nodes that conform  $\bar{F}_{15}$  and the outcome node  $Y_{15}$  (Härkönen, Bernardi, and Boertien 2017). Both of these hypothesis consider, to some extent, how each factor  $V_t$  causes dynamic confounding (Lee and McLanahan 2015).

The DAG also illustrates the importance of understanding which factors determine the realization of a particular instability trajectory. Knowledge and observation of these factors is necessary to obtain a consistent and unbiased estimation of the causal effect of exposure to family instability. The vector  $\bar{F}_{15}$  could be made up of specific family structures experienced at each time-point – which are known as the alphabet in the sequence analysis literature (Gauthier et al. 2010, 7) – but they can be thought of as binary indicators of any kind of change in family structure ( $F_t \in \{0, 1\}$ ), marking episodes of family changes. In this framework, two changes in family structure can take, among many others, for example, the following forms:  $\{0, 1, 1, 0, 0, 0\}$ ,  $\{0, 0, 1, 1, 0, 0\}$ ,  $\{0, 0, 0, 0, 1, 1\}$ , or  $\{0, 1, 0, 0, 1, 0\}$ , thus marking the different time points when changes in family structure took place.

Estimating the difference between adolescent problem behavior measures for children following different family trajectories, such as  $Y_{15}^{(0,0,0,0,0,0)}$  and  $Y_{15}^{(0,0,0,0,1,1)}$  conditional on some time-invariant or varying covariates, does not warrant a causal interpretation. As explained above, children who were exposed to family changes will likely differ on observed and unobserved characteristics from children exposed to alternative trajectories. The DAG in Figure 8 shows the many backdoor paths that remain open or upon up when inappropriately adjusting for them, employing conventional statistical methods. The gray and blue arrows show the paths will bias the estimates of

causal effects (i.e. the red arrows) one is interested in obtaining, when using methods such as ordinary least squares. Any effects of family instability up to any given time point after baseline ( $\bar{F}_t$ ) will be biased when adjusting only for baseline covariates ( $W_0$ ) that do not vary over time. But will also be biased by naively adjusting by the history of time-varying confounders ( $\bar{V}_t$ ) through regression (e.g. generalized linear mixed models), stratification or matching, or by simply neglecting unobserved pathways. The trajectories captured by the time-varying confounders  $\bar{V}_t$ , and the family instability trajectories  $\bar{F}_t$  are interlocked in a feedback chain of cause and effect.

The overall aim of this paper is to find a more adequate adjustment for the dynamic confounding factors affecting the sequence of family transitions. The effects of interest can be estimated by a doubly robust MSM, estimated employing IPTW. MSM serve the goal of modeling the marginal distribution of the unobserved potential outcomes in data when time-dependent exposure, time-dependent confounding, and treatment-confounder feedback effects are present (Hernan and Robbins 2020, 247–55). MSM will provide unbiased estimates in the case of dynamic confounding and treatment-confounder dependencies, under the assumption of having reached sequential exchangeability, positivity, and consistency. However, sequential exchangeability relies on assuming that no further unmeasured confounders remain. But as I argue in the discussion section, this assumption is strong in light of neighborhood effects (Sharkey and Faber 2014), parental wealth (Hällsten and Pfeffer 2017) and marital sorting in social networks (Downey and Condrón 2016), which likely affect family changes and children’s wellbeing, to name a few of the potential time-varying confounders which remain partly or fully unobserved in this study. Furthermore, the consistency assumption only holds if models of the selection into family instability (i.e. the propensity models) are correctly specified. To avoid misspecification due to incomplete knowledge of the determinants of a family

instability trajectory, a doubly robust version is used to reach consistency.

The estimation of MSM is done in two steps. In the first step, probability weights obtained from IPTW – in this case employing BART (J. L. Hill 2011) and improving the specification of selection models – are used to reweigh the sample of children and obtain a pseudo-population on which we can estimate effects on counterfactual unobservable outcomes. I used stabilized weights ( $sw_i^t$ ) to account for the low probabilities of specific trajectories of family instability in the sample up to time  $t = 15$ , which is when children in the cohort have reached adolescence. The weights  $sw_i^{15} = \frac{h(f_t|\bar{f}_{t-j},w_o)}{h(f_t|\bar{f}_{t-j},\bar{v}_{t-1},w_o)}$  for a sample of  $i = 1, \dots, N$  children,  $j = 1, 3, 5, 9, 15$  denoting the previous waves of the study. Here the function  $h()$  corresponds to the conditional distribution of the exposure at time point  $t$ , also known as a propensity score model. I employ BART to estimate these conditional distributions for the event of changing family structure during childhood between follow-up surveys  $t$  and  $t - 1$ . BART is a regression method that can flexibly fit non-linear outcomes in high-dimensional inference problems (J. Hill, Linero, and Murray 2020), when many covariates are available. One advantage of this method, which can further be used for causal inference in observational studies, is that it has the ability to capture relevant interactions and non-linearities among the covariates (Tan and Roy 2019). In this paper, I use BART to obtain the estimates of propensity scores based on a causal inference model for the effect of changes in family structure at each time point on problem behavior among fifteen-year-old children.

The propensity scores obtained from these models are then used in the estimation of the MSM. For the probabilities in the numerator, I used prior changes in family structure and baseline family structure, together with baseline confounders that do not change over time. For the denominator, information on the history of previous changes in family structure, all time-varying confounders up to  $t - 1$ , and all baseline confounders were

used. In both cases, I used BART to estimate the propensity scores for children experiencing a family change at each time-point. To account for loss to follow-up, an additional censoring model with baseline covariates was estimated to compute weights that account for selective drop out (i.e. avoiding sample selection bias). I obtained the predicted probabilities of changing family structures between each follow-up survey and censoring, and combined them to form the stabilized weights obtaining  $sw_i^{15} = sw_i^1 \times sw_i^3 \times sw_i^5 \times sw_i^9 \times c_i^{15}$  where the superscript 15 denotes the time-point at which effects on children's problem behavior are estimated.

In the second step of the estimation of MSM, a regression model for the outcome variables in which baseline covariates determining the outcome are included. Thus, a weighted regression model adjusting for further baseline covariates will be doubly robust, and will have higher chances of capturing the effect if the adjusted outcome regression model or the propensity score model, or both, are correctly specified (Kang, Schafer, et al. 2007). A first model is estimated for the life course cumulative effects model, denoted here as

$$\mathbb{E}(Y_t^{\bar{f}_t}) = \beta_0 + \beta_1 cum(\bar{f}_t) + \beta_2 W_0 \quad (1)$$

where  $cum(\bar{f}_t) = \sum_{t=0}^{\tau} f_t$  is the sum of instability episodes within the trajectories observed up to a specific time point  $\tau$ , and the estimates of  $\beta_1$  corresponds to the effect of those trajectories on a counterfactual outcome  $Y_t^{\bar{f}_t}$ . A second model 2 is also estimated, which represent a more complex, though high demanding, fully interacted or saturated life course model

including all potential interactions between the exposures of interest

$$\begin{aligned}\mathbb{E}(Y_t^{\bar{f}_t}) = & \theta_0 + \theta_1 F_1 + \theta_2 F_3 + \theta_3 F_5 + \theta_4 F_9 + \theta_5 F_{15} \\ & + \theta_7 F_1 F_3 + \theta_8 F_3 F_5 + \theta_9 F_3 F_9 + \theta_{10} F_3 F_{15} + \theta_{11} F_1 F_3 F_5 \\ & + \dots + \theta_{32} F_1 F_3 F_5 F_9 F_{15} + \theta_{33} W_0\end{aligned}\quad (2)$$

Whereas Model 1 is testing the cumulative effects of family instability, irrespective of their timing, Model 2 simultaneously compares all potential trajectories, directly encoding the timing, which can be tested by looking at the estimates of each coefficient capturing specific interactions. I estimate linear regression for the outcome models, comparing various instability trajectories that differentiate the number and timing of the effects. Positive values of estimates in these regression models on the re-weighted sample imply that the outcome is higher for children exposed to unstable family trajectories. I considered a standardized or rescaled version of the score on each of the problem behavior scales so that parameter estimates may be interpreted as standardized coefficients and compared across time-points. The average percentage of missing information in the analytical variables was around 16%. To address bias due to missing information, I used multiple imputation by chained equations (van Buuren and Groothuis-Oudshoorn 2011) assuming missing at random. The imputation model used the CART algorithm (Burgette and Reiter 2010) to find the best set of predictors among the analytical variables to impute the data using all information available across waves, with 20 iterations per imputation.

## Data

In the FFCWS, researchers have followed a cohort of American children for over twenty years. This cohort was born between the years 1998 and 2000 in U.S. cities with a population greater than 200,000 (Reichman et al. 2001). The study is based on a probabilistic sample with a complex design in which

cities, hospitals and beds in randomly selected hospitals were sampled. The FFCWS oversampled births to unmarried parents, a group of parents that experiences more frequent family instability but high attrition in previous studies. For an overview of the data response rates, sample weights and sampling designs see FFCWS (2019) and Kennedy and Gelman (2018).

Women and men of certain characteristics might self-select into lone-parenthood or be more prone to divorce and to re-partner, or may be exposed to restricted marriage markets, which limit their chances of finding a partner (see Cohen and Pepin 2018; Lichter, Price, and Swigert 2019; and Wilson 1987). Some of those characteristics can also impact children's outcomes in systematic ways. For this reason I included a rich set of baseline characteristics observed at birth which are in line with previous research (F. Bernardi et al. 2013).

***Time-invariant and time-varying adjustment variables*** I included two sets of characteristics that may confound and bias the causal estimates, one corresponding to the time-constant or baseline characteristics; and another one to the time-varying characteristics that act as both mediators and confounders. Table 2 provides an overview of these variables, their operationalization, as well as the time-points at which they were captured in the FFCWS. There is little knowledge about which specific factors may affect the probability of experiencing a given family instability trajectory (Härkönen 2014; Saint-Jacques et al. 2017). For this reason, I selected a broad set of time-dependent factors that are likely contributing to changes in partnership trajectories of mothers in causal ways, based on previous research and the family instability hypothesis. This selection of variables is in line with the work of Lee and McLanahan (2015) and others, but I added many other variables that were not considered in any previous study in a dynamic fashion. The time-invariant and time-varying variables sets are,

however, partial, meaning that other characteristics that confound the association remain unobserved. I make use of the child's mother and father responses to the questions, which were obtained at baseline and all other available follow-ups when children were one, three, five, nine and fifteen years old.

***Measures of child's behavioral problems*** Child behavior problems, both internalizing and externalizing ones, were assessed at the third, fifth, ninth and fifteenth follow-ups, but in this paper I focus on the problem behavior scores among 15-year-olds. These different scores are based on different subsets of questions from the Child Behavior Checklist 2-3 (Koot et al. 1997). Maternal ratings on a three level scale indicating whether a given item was not true (0); sometimes or somewhat true (1); or very true or often true (2) of her child make up the total score on both the internalizing and externalizing dimensions. Table 3 shows the different items that compose the indicators of problem behavior at age fifteen years old.

***Changes in family structure*** The effects of family instability are usually examined by considering two different dimensions of family dynamics: First, the relationship status between the biological parents of the child; and second, the presence or absence of the biological or social father. The operationalization of childhood family structure at each time-point follows Raley et al. (2019). I used the mother's relationship with the biological father of the child at each wave together with information on the relationship of the child to the people living in her household:

- If the mother was married to or cohabiting with the biological father of her child and lived in a household in which her partner was declared to be present, then I assumed these were the "traditional" two biological nuclear families ( $T$ ).

- If she was married to or cohabiting with her partner, but did not live in a household with her partner present, then these were classified as living apart together ( $L$ ) families.
- If the mother was not in a relationship with the father of the child (i.e. in a visiting relationship, they were friends, never or hardly talk, were divorced, separated or widowed, or had no relationship), and the mother did not live in a household with a partner present, then these were single-mother families ( $S$ ).
- Lastly, if the mother was not in a relationship with the father of the child, but lived in a household with a partner present, then these were considered as blended, complex or stepfamilies ( $B$ ).

After classifying childhood family structure in any of these four types ( $f_t \in \{T, S, B, L\}$ ), and at each time point, I operationalized family instability as a dichotomous indicator variable capturing any change in the family structure. For the purposes of this study, the family instability indicator, thus, takes the value of one when the family structure at time  $t$  differs from the one observed at  $t - 1$ , the previous observation period, and zero otherwise, such that instead of specific family structures, at each time point we have  $f_t \in \{0, 1\}$ . Although this definition does not distinguish between the different types of changes (Cavanagh and Huston 2008), I am able to capture the most recently occurred changes in family structure. Related to this, and as in previous research (Lee and McLanahan 2015), between survey waves, and especially between those waves spread over multiple years, I did not capture possible further changes in family structure which might have occurred. Therefore, the indicator is probably a conservative estimate of the real number of family transitions.

## Results

Table 4 presents descriptive characteristics of the FFCWS sample at baseline by the type of family structure at birth, for which all further analyses adjust for. Substantial differences can be observed between children born to married or cohabiting parents v. those born to single-mothers or in blended, complex or stepparent family configurations – the children born to LAT family structures are in contrast more similar to the married/cohabiting parents. The “Fragile families,” those that were not married at the child’s day of birth, are more prone to experience family instability, are predominantly formed by young and low-educated parents, and mostly of non-Hispanic Black and Hispanic racial/ethnic backgrounds. They are more likely to live in with extended or unrelated kin, and are more affected by poverty than the married two-parent families. Moreover, different exposures to changes in family structure over time can have different effects at various later outcomes. Figure 9 shows the average total problem behavior score evaluated at four different ages according to the different exposure histories of changes in family structure. One first remarkable result is that the differences in problem behavior between the distinct exposure histories are not large. In fact, there is no easily discernible pattern between more episodes of family instability, or specific timings, and a higher score on overall problem behavior. The most stable trajectory at each time point, however, has the smallest average score. For the remaining of the paper, I focus on problem behavior, and its internalizing and externalizing dimensions, among adolescents, shown in panel D in Figure 9.

The results of the MSM fitted to the life course cumulative instability model are shown in Figure 10. These are the effects of different number of changes, irrespective of the timing of the family changes, compared to the reference category of no changes in family structure, i.e., the stable families. The

conditional effects, unadjusted and adjusting only for baseline time-invariant covariates, are shown next to the causal estimates obtained from the doubly robust MSM based on IPTW and adjustment for time-invariant covariates. In general, estimates of the unadjusted effect, which for most outcomes show the well-known, moderate negative association between family instability and children’s wellbeing, are substantially reduced by the inclusion of baseline confounders; and slightly further reduced by consideration of the dynamic interplay of time-dependent confounders, with a few noticeable exceptions.

As suggested by Figure 10, most of the action is on the externalizing problem dimension, given that internalizing problems do not respond to more changes in family structure – seen in that the curve is mostly flat. In contrast, externalizing problems tend to increase with the number of changes, but not linearly. After three changes, when the effect is largest, though still of a small magnitude, the MSM estimates suggest that exposure to further changes in family structure have a smaller effect. The effect for three changes in family structure experienced during childhood likely correspond to a common experience: The parents’ separation or divorce, followed by a repartnering, and a subsequent separation from the new partner. However, three changes can in principle capture other types of transitions as well.

Though all confidence intervals in Panel C Figure 10 contain the value zero, the effect of three changes on externalizing problem behavior are remarkably consistent. The adjustment for time-dependent confounding does not reduce the size of this estimate much more, though it slightly increases the standard error. This result, in contrast to all other comparisons being made with the life course cumulative model, suggest the presence of a small negative effect of at least three episodes of family instability on the externalizing problem behavior of teenagers, which goes in line with

previous results (Lee and McLanahan 2015; Rodriguez S. 2021). However, the estimates for further changes, four and five, are substantially reduced, with children experienced up to five changes showing lower problem behavior than children in stable families, which speaks against the family instability hypothesis (Hadfield et al. 2018).

To explore matters further, Figure 11 shows the moderating effects in the cumulative model by racial-ethnic background and the child's gender assigned at birth, based on the doubly robust MSM employing IPTW. Panel A shows the moderating effect of racial-ethnic background, which suggests that White children are on average more affected than Hispanic children for all number of changes, whereas no sign of moderation for Black children with the exception of the effect of five changes. Panel B, in turn, explores the moderation by the child's gender assigned at birth, which suggests that boys exposed to five family changes are more affected than girls in both externalizing and internalizing dimensions, whereas no moderation effect is observed for smaller number of changes.

These findings already suggest important forms of heterogeneity in the effect of cumulative changes or number of family transitions, a heterogeneity that can hardly be captured by the family instability hypothesis. What could possibly explain these findings? One possibility is that the specific family life course trajectories experienced by children have effects that do not correspond to major theoretical expectations, hence the relevance of evaluating whether the timing makes a difference.

Figure 12 is based on the more flexible, fully interacted doubly robust MSM, which tests whether the life course pathways model can reveal more nuances in the claim of negative effects. This model contains all possible interactions between the different time points at which children experienced a family transition, and can in principle show the effect of specific joint exposure

histories. As already hinted at in the descriptive results, and zooming in on the internalizing and externalizing dimensions, the effects of these different exposure trajectories are far from straightforward. The same total number of family transitions can have both positive and negative effects. However, it is impossible to discern any clear pattern regarding the effects of timing or number of changes from this model, especially considering the relatively small sample size for this kind of analysis. One highlight of this figure, however, is that trajectories with less number of family transitions are associated with a higher effect than trajectories with more family transitions, and sometimes even of an opposite sign to the theoretically expected. Moderating effects in the pathways model cannot be reliably estimated given the relatively small sample size (i.e., not enough units to compare specific trajectories by subgroups defined by the child's gender or racial-ethnic background).

## **Discussion**

The results of this paper suggest that the association between family life course trajectories with adolescent problem behavior is far more complex than previously thought of. This higher complexity has multiple sources, as seen in the moderation of some effects, suggestive of cumulative changes in family structures affecting certain subgroups (Lee and McLanahan 2015), but mainly in that family life course trajectories may not follow simple monotonic rules and, more generally, may not correspond to the causal models often hypothesized in the literature. In summary, this re-examination of the family instability hypothesis suggests that changes in family structure may not be as detrimental to problem behavior among adolescents. The data here analyzed further suggest the family instability hypothesis portrays family dynamics in an overly simplistic manner, disregarding the potential for highly complex effects of family dynamics and

its associations with children's problem behavior.

A life-course reading of the family instability hypothesis and dynamic confounding reveals how untangled causes and consequences can get in the study of family dynamics (L. Bernardi, Huinink, and Settersten Jr 2019). Past research has argued that family instability is a dynamic phenomenon with multiple negative consequences for the wellbeing of children (Cavanagh and Fomby 2019; Lee and McLanahan 2015), but an account of the sequencing of events – including both reinforcing and counteracting processes – set in motion in the life-course of children by a change in family structure shows this might not be the case. These intermediate processes may attenuate to a considerable extent the presumed negative impacts of family instability; thus suggesting there might not be negative effects after all. This in turn would imply that various indirect feedback mechanisms may be as relevant for children's wellbeing, as the initial triggering family change.

A study of family life course trajectories defined by specific family structures experienced by children, e.g., living with two parents, living with one parent, a stepparent, a second stepparent, or even considering further household dimensions, such as coresidence with other relatives, etc., suggests itself as the next step in understanding why the number of family changes can have such varied effects.

However, such an analysis would require a far larger number of observations followed from birth onward because of two reasons: 1) there is increasing variability as children get older and the sequences of family life course are quite complex from a statistical modeling point of view; and 2) by the time children are nine years old, there are very few children following specific family trajectories, which limits the extent to which we can establish statistical comparisons for more detailed family life course.

Given selection on relevant unobservable variables and the consequent likely

misspecification of the propensity score models estimating the probabilities for family changes at each time point, the IPTW estimates may still be upwardly biased. For example, no study on the effects of changes in family structure has considered the effects of neighborhoods, social networks, yet the decision to marry, to have children, and to stay married to a specific partner, may be affected by these. Both of these confounding factors may affect the stability of families because the cultural capital and the economic capital or patrimony acquired through marriage may be at stake in the case of a divorce. Individuals in specific social circles with high economic capital marry “because of their capital” and stay together “for their capital” (Bourdieu 1993) [p. 35]. On the opposite, lack of wealth or low socio-occupational attainment may decrease the stability of family among those who had little to gain from marriage to begin with, and therefore little to lose from its dissolution; while at the same time, wealth and socio-occupational class are important determinants of children’s wellbeing (Downey and Condron 2016; Hällsten and Pfeffer 2017; and for other contextual factors see Fomby and Mollborn 2017). Future research should address this as well as other potential pathways that affect the dynamic feedback effects between the set of time-dependent confounders  $\bar{V}_t$  and family instability  $\bar{F}_t$ .

Measurement error may induce further bias thus far not considered in this paper. For example, psychological wellbeing variables such as those used for internalizing and externalizing problems, are unobserved constructs measured with error. Of special concern in this regard are the scales built from maternal ratings. Family transitions may make some mothers more keen observers of their child’s behavior than others. If family instability affects how mothers rate this behavior, artificially a statistical association is created between changes in family structure and the score obtained in a ratings checklist. So far, the literature has not discussed this in relation to

the estimation of family instability effects. There is, however, little that can be done to address this issue with current statistical tools, under not so restrictive assumptions, besides trying to device better measurement instruments.

The life-course of children who experience changes, even repeated changes, in family structure tell a more complex story. A life-course perspective reveals that family instability should be considered as a dynamic concept, rather than a static one, as suggested by Wu (1996). Doing so moves forward the debate in the direction proposed by Seltzer (2019) for demographic studies of families, by better portraying contemporary dynamics of family life, but it follows that more flexible models are needed for understanding the complex effects family instability can have on children's wellbeing. Therefore, children negatively affected by instability in family structure may be only those for which the triggering negative event of family instability led primarily to reinforcing or counteracting mechanisms, which may be caused by further unobserved sources of heterogeneity. The resulting chain of negative events that result from changes in family structure can turn into a cumulative disadvantage trajectory in multiple dimensions of the life-course, but not all children, and, in fact, most children, experience both reinforcing and counteracting events in their life-course after changes in family structure take place, which may lead to a positive balance after the complete series of events is accounted for.

Future research ought to be aware of the role of family privilege in family theory and research as suggested by Hadfield, Ungar, and Nixon (2018). Conceiving stable families, especially those conformed by the two biological parents of the child, as the benchmark to which other family arrangements ought to be compared to, reminds us of the ethnocentric perspective of White middle-class researchers who insist that children “fare better” under—demographically speaking—living arrangements most similar to their own.

More sound understandings of family dynamics ought to make room for people's agency in their choices of family life, as these are affected by the structural constraints in which they try to navigate the demands of work and family (Damaske, Bratter, and Frech 2017). Family life is equally determined by potential alternative sources of strength and resilience which may be found in alternative family structures as well (Orthner, Jones-Sanpei, and Williamson 2003). Against the hypothesized harmful psychological effects of "dysfunctional" families or absent relationship figures in children's lives, this paper highlights the extent to which socioeconomic trajectories of families (i.e. employment, income, housing and household trajectories) matter in explaining children's wellbeing.

If the results of this paper seem puzzling at first sight, given the strong theoretical expectations of negative effects, it is only because research has been disproportionately focused on the reinforcing mechanisms that would follow a family transition such as divorce. Family research should also reconsider the role played by counteracting mechanisms that are set in place once a family transition takes place (Kelly and Emery 2003). Many of these mechanisms, as explained in this paper, remain under-theorized. For example, help from extended family members may only become available after the father leaves the household. Future research should focus on establishing the strength of both reinforcing and the counteracting mechanisms. The results of this paper suggest that these tend to, on average, balance out whatever negative effect would supposedly follow from a family transition. However, this research must find how those processes moderate the outcome as to generate a net null effect of family instability on children's wellbeing. Methods such as structural nested models may provide evidence for this but this task remains outside the scope of this paper.

## Tables and Figures

Table 2: Description of variables used in chapter 3

Variables	Min	Max	N	Description	Waves
Child's verbal ability (PPVT)	40	137	964	Peabody Vocabulary Test (PPVT/TVIP) score	3rd, 5th, 9th years
Child's combined GPA	1.00	4.00	1199	It includes the subjects of English, Math, History or Social Studies, and Science	15th year
Ever failed a class	1	2	1422	Ever failed a class in school? (yes/no)	15th year
Emotionality	3	15	1831	Emotionality - EAS Temperament Survey for Children: Parental Ratings - the tendency to become aroused easily and intensely, broad measure of distress in the very young infant, associated to fear and anger tendencies in the older child.	1st year
Shyness	3	15	1827	Shyness - EAS Temperament Survey for Children - a tendency towards inhibition and awkward behavior in the young child	1st year
Child behavior problems	0	80	1156	Child Behavior Problems (CBCL) including both internalizing and externalizing problems	3rd, 5th, 9th, and 15th years
Trouble at school scale	0	8	1411	Trouble at School in four areas getting along with teachers, paying attention in school, getting homework done, and getting along with other students (0 = Never to 4 = Every day)	15th year
Ever suspended from school	1	2	1423	Ever been suspended or expelled in past 2 years? (yes/no)	15th year
Diagnosed with asthma	1	2	1628	Has a health care professional ever told you child has asthma? (yes/no)	3rd, 5th, 9th, and 15th years

Table 2: Description of variables used in chapter 3 (*continued*)

Variables	Min	Max	N	Description	Waves
Overweight	1	2	981	Child Body Mass Index standardized by age and gender, binary indicator if child's BMI is higher than the 85th percentile (yes/no)	3rd, 5th, 9th, and 15th years
Use of substances	1	2	1452	Ever smoked an entire cigarette? Drank alcohol more than two times without parents? Tried marijuana? Tried other illegal drugs besides marijuana?	15th year
Child's gender	0	1	2055	Gender assigned at birth (boy/girl)	Baseline
Low birth weight	0	1	1999	Was the child diagnosed with low birth weight? (yes/no)	Baseline
Mother drank alcohol during pregnancy	1	5	2049	During the pregnancy, how often did you drink alcohol? (1=Everyday to 5=Never)	Baseline
Mother smoked during pregnancy	1	4	2050	During the pregnancy, how many cigarettes did you smoke? (1=more than packages per day - 4=None)	Baseline
Mother took drugs during pregnancy	1	5	2051	During the pregnancy, how often did you use drugs? (1=Everyday to 5=Never)	Baseline
Father's last name on birth certificate	0	1	2042	Will the baby (babies) have the father's last name? (yes/no)	Baseline
Mother's age at child's birth	15	43	2054	Mother's age (years)	Baseline
Father's age at child's birth	16	53	1858	Father's age (years)	Baseline

Table 2: Description of variables used in chapter 3 (*continued*)

Variables	Min	Max	N	Description	Waves
Mother's education at child's birth	1	4	2053	Mother's education (1=less than highschool, 2=highschool or equivalent, 3=some college, technical education, or 4=college or graduate)	Baseline
Mother's verbal ability (PPVT)	40	139	929	Mother's or primary care giver PPVT - Standardized score	3rd year
Father's education at child's birth	1	4	2046	Father's education (1=less than highschool, 2=highschool or equivalent, 3=some college, technical education, or 4=college or graduate)	Baseline
Mother's race	1	4	2051	Mother's race 1= white, non-hispanic; 2 black, non-hispanic; 3=hispanic; 4=other	Baseline
Father's race	1	4	2054	Father's race 1= white, non-hispanic; 2 black, non-hispanic; 3=hispanic; 4=other	Baseline
Mother is US citizen	0	1	2053	Was the mother born in the U.S.? (yes/no)	Baseline
Father is US citizen	0	1	1856	Was the father born in the U.S.? (yes/no)	Baseline
Mother's religiosity	0	1	2052	How often does mother attend religious services? (at least once a year vs. never)	Baseline
Father's religiosity	0	1	1857	How often does father attend religious services? (at least once a year vs. never)	Baseline
Mother lived with both her parents	0	1	2035	Was the mother living with both of her biological parents at age 15? (yes/no)	Baseline

Table 2: Description of variables used in chapter 3 (*continued*)

Variables	Min	Max	N	Description	Waves
Father lived with both his parents	0	1	1847	Was the father living with both of her biological parents at age 15? (yes/no)	Baseline
Mother thought about abortion	0	1	2051	When mother found out she was pregnant, did she think about having an abortion? or the father suggested her to have an abortion?	Baseline
Father thought about abortion	0	1	1852	When father found out the biological mother was pregnant, did he think about her having an abortion?	Baseline
Mother's overall health	0	1	2054	How is the mother's health? (1=Great to 5=Poor)	Baseline, 3rd, 5th, 9th, and 15th years
Mother-father relationship quality	1	3	2040	After mother found out she was pregnant, how did mother's relationship with biological father change? (better, worse, same) And after baseline, how is mother's relationship with child's father? (1=excellent to 5=Very bad)	Baseline, 3rd, 5th, 9th, and 15th years
Mother with alcohol/drug problems	0	1	2052	In last year, have alcohol/drugs interfered with mother's work/relationships? (yes/no)	Baseline, 3rd, 5th, 9th, and 15th years
Father with alcohol/drug problems	0	1	1855	In last year, have alcohol/drugs interfered with father's work/relationships? (yes/no)	Baseline, 3rd, 5th, 9th, and 15th years
Father has been in jail	0	1	2053	Both mother and father report that father was in jail at each interview (yes/no)	Baseline, 3rd, 5th, 9th, and 15th years
Violence against the mother	0	1	1955	Frequency that father hit or slaps mother when he is angry, insults or criticizes her? (if this ever happened one, else zero)	Baseline, 3rd, 5th, 9th, and 15th years

Table 2: Description of variables used in chapter 3 (*continued*)

Variables	Min	Max	N	Description	Waves
Welfare/food stamps/TANF recipient	0	1	2039	In last year, did the mother have income from public assistance/welfare/food stamps/TANF? (yes,no)	Baseline, 3rd, 5th, 9th, and 15th years
Financial assistance from other family members	0	1	2050	Have you receive financial support from anyone besides biological father?	Baseline, 3rd, 5th, 9th, and 15th years
Father's socio-occupational category	1	7	1848	What sort of work does/did father do in his current/most recent job? (white collar, high skill; services, high skill; manual blue collar; other low skill; self-employed; unemployed; or out of the labor force)	Baseline, 3rd, 5th, 9th, and 15th years
Mother's socio-occupational category	0	1	2055	What sort of work does/did mother do in his current/most recent job? (white collar, high skill; services, high skill; manual blue collar; other low skill; self-employed; unemployed; or out of the labor force)	Baseline, 3rd, 5th, 9th, and 15th years
Neighborhood violence level	1	4	2045	How safe are the streets around your home at night or frequency of gang activity? (1= Very safe to 5=Very unsafe)	Baseline, 3rd, 5th, 9th, and 15th years
Mother has moved from previous house	0	1	1583	Has the mother moved houses since child was born or since last interview? (yes/no)	Baseline, 3rd, 5th, 9th, and 15th years
Mother lives in a rented house/apartment	0	1	2038	Is the home/apartment were mother currently resides owned/rented?	Baseline, 3rd, 5th, 9th, and 15th years

Table 2: Description of variables used in chapter 3 (*continued*)

Variables	Min	Max	N	Description	Waves
Non-Nuclear family structure	0	1	2055	A synthetic indicator created from household members information (dichotomous indicator if non-nuclear family structure is present)	Baseline, 3rd, 5th, 9th, and 15th years
Child's siblings live in the same household	0	1	2055	A synthetic indicator created from household members information (dichotomous indicator if at least one child's sibling is present)	Baseline, 3rd, 5th, 9th, and 15th years
Housing wealth	-315002	5000000	1041	Net housing wealth (difference between the value the house could be sold minus what is owed to the bank)	Baseline, 3rd, 5th, 9th, and 15th years
Equivalized household income	0.0000	94575.5320	2041	Household income combining all sources of income, divided by the square root of the household size	Baseline, 3rd, 5th, 9th, and 15th years
Poverty categories	1	5	2055	Poverty categories constructed by FFCWB based on mother's household income/poverty threshold ratio (1 = 0-49% to 5 = 300%+)	Baseline, 3rd, 5th, 9th, and 15th years
Mother's depression	0	1	1839	Mother meets depression criteria (liberal) at one-year (based on the CIDI questionnaire)	3rd, 5th, 9th, and 15th years
Child's age (at second wave)	9	30	1837	Baby's age at time of mother's one-year interview	3rd, 5th, 9th, and 15th years
Multipartner fertility indicator	0	1	1833	Mother has children by man other than the biological father of child	3rd, 5th, 9th, and 15th years

*Note:*

Own elaboration.

Table 3: Average score on items and subscales of problem behavior at fifteen-year follow-up and descriptives by gender and racial-ethnic background

Dimension	Question or item	Total	Girls	Boys	White	Black	Hispanic
Withdrawn	Youth is underactive, slow moving, or lacks energy	0.32	0.33	0.30	0.34	0.30	0.34
	Youth is unhappy, sad, or depressed	0.28	0.31	0.26	0.35	0.27	0.25
Anxiety/Depression	Youth cries a lot	0.13	0.22	0.07	0.19	0.12	0.13
	Youth feels worthless or inferior	0.16	0.17	0.18	0.27	0.14	0.15
	Youth is nervous, highstrung, or tense	0.30	0.32	0.30	0.42	0.27	0.30
	Youth is too fearful or anxious	0.26	0.27	0.25	0.31	0.22	0.26
	Youth feels too guilty	0.14	0.15	0.15	0.17	0.14	0.14
	Youth worries	0.48	0.52	0.49	0.67	0.43	0.47
Internalizing	Subtotal	2.04	2.28	1.98	2.70	1.87	2.02
Aggressive	Youth argues a lot	0.47	0.50	0.48	0.55	0.44	0.50
	Youth is cruel, bullies, or shows meanness to others	0.19	0.18	0.20	0.15	0.22	0.16
	Youth destroys things belonging to family or others	0.10	0.08	0.14	0.09	0.13	0.09
	Youth is disobedient at home	0.42	0.39	0.47	0.41	0.45	0.41
	Youth is disobedient at school	0.29	0.23	0.33	0.17	0.40	0.21
	Youth gets in many fights	0.10	0.11	0.10	0.06	0.13	0.08
	Youth physically attacks people	0.05	0.03	0.06	0.02	0.06	0.05
	Youth is stubborn, sullen, or irritable	0.62	0.64	0.63	0.72	0.65	0.50
	Youth has temper tantrums or a hot temper	0.39	0.38	0.41	0.35	0.43	0.38
	Youth threatens people	0.08	0.06	0.10	0.07	0.10	0.07
	Youth is unusually loud	0.33	0.38	0.29	0.30	0.38	0.28
	Youth doesn't seem to feel guilty after misbehaving	0.33	0.30	0.36	0.28	0.39	0.30
Rule Breaking	Youth hangs around with others who get in trouble	0.21	0.17	0.23	0.15	0.28	0.15
	Youth lies or cheats	0.32	0.27	0.37	0.26	0.39	0.27
	Youth runs away from home	0.04	0.04	0.04	0.03	0.04	0.03
	Youth sets fires	0.01	0.00	0.02	0.01	0.01	0.01
	Youth steals at home	0.07	0.05	0.07	0.04	0.08	0.05
	Youth steals outside the home	0.04	0.03	0.05	0.03	0.06	0.01
	Youth swears or uses obscene language	0.37	0.33	0.42	0.44	0.38	0.34
	Youth vandalizes	0.03	0.03	0.04	0.03	0.05	0.04
Externalizing	Subtotal	4.39	4.16	4.70	4.11	5.02	3.80
Behavior Problems	Total	10.31	10.35	10.65	10.75	10.97	9.55

*Note:*

Own elaboration. All items are on a scale from 0 (not true), 1 (sometimes or somewhat true) to 2 (very true or often true)

Table 4: Descriptive Statistics for Analytical Variables in the FFCWS Initial Sample by Family Structure at Birth

		Married or Cohabiting two parents	Lone-parent	Complex, step or blended family	LAT	Missing
N	Count	1979	1286	28	148	1
Mother's age	Median	25	22	23.5	25	22
Father's age	Median	28	24	26	27	-
Mother's race (%)	White, non-hispanic	32.4	12.9	28.6	19.6	0
	Black, non-hispanic	30.2	58.5	57.1	43.9	0
	Hispanic	32.1	25.7	14.3	33.8	100
	Other	5.1	3	0	2.7	0
	Missing	0.2	0	0	0	0
Father's race (%)	White, non-hispanic	30.2	8.9	10.7	15.5	0
	Black, non-hispanic	31.7	61.4	57.1	43.9	0
	Hispanic	32.8	24.8	21.4	34.5	100
	Other	5.3	3.2	7.1	6.1	0
	Missing	0.1	1.7	3.6	0	0
Mother born in U.S. (%)	No	19.6	10.7	3.6	26.4	0
	Yes	80.3	89	96.4	73.6	100

Table 4: Descriptive Statistics for Analytical Variables in the FFCWS Initial Sample by Family Structure at Birth (*continued*)

		Married or Cohabiting two parents	Lone-parent	Complex, step or blended family	LAT	Missing
Father born in U.S. (%)	Missing	0.1	0.3	0	0	0
	No	18.2	5.2	3.6	22.3	0
	Yes	73.1	55.8	57.1	52	0
Mother's religiosity (%)	Missing	8.6	39	39.3	25.7	100
	Seldom or never	39.8	43.2	21.4	30.4	100
	At least once a year	60.2	56.6	78.6	68.2	0
Father's religiosity (%)	Missing	0.1	0.2	0	1.4	0
	Seldom or never	40.9	30.1	28.6	29.1	0
	At least once a year	50.5	30.9	32.1	45.3	0
Mother thought about abortion (%)	Missing	8.6	39	39.3	25.7	100
	No	79.8	50.3	53.6	77.7	0
	Yes	20	46.6	39.3	21.6	0
	Missing	0.3	3.1	7.1	0.7	100

Table 4: Descriptive Statistics for Analytical Variables in the FFCWS Initial Sample by Family Structure at Birth (*continued*)

		Married or Cohabiting two parents	Lone-parent	Complex, step or blended family	LAT	Missing
Father suggested abortion (%)	No	78.6	43.5	42.9	64.9	0
	Yes	12.6	17.3	17.9	9.5	0
	Missing	8.8	39.3	39.3	25.7	100
Mother's education level (%)	Less than Highschool	29.3	40.7	32.1	34.5	0
	Highschool or GED	27.6	33.2	28.6	29.7	100
	Some college or tech.	26	23.1	28.6	22.3	0
	College or Graduate	17.1	3	10.7	13.5	0
	Missing	0.1	0.1	0	0	0
Father's education level (%)	Less than Highschool	30.5	35.9	39.3	36.5	0
	Highschool or GED	28.8	34.2	25	31.1	0
	Some college or tech.	24.4	16.7	17.9	20.3	0
	College or Graduate	16	3.3	0	10.8	0
	Missing	0.4	9.9	17.9	1.4	100

Table 4: Descriptive Statistics for Analytical Variables in the FFCWS Initial Sample by Family Structure at Birth (*continued*)

		Married or Cohabiting two parents	Lone-parent	Complex, step or blended family	LAT	Missing
Mother lived with both parents growing up (%)	No	49	65.9	71.4	49.3	100
	Yes	50.3	33.4	28.6	45.9	0
	Missing	0.7	0.8	0	4.7	0
Father lived with both parents growing up (%)	No	43.2	39.4	39.3	37.2	0
	Yes	48	21.5	21.4	35.1	0
	Missing	8.9	39	39.3	27.7	100
Mother's health	Good	93.2	91.7	92.9	93.9	100
	Fair or poor	6.7	7.9	7.1	6.1	0
	Missing	0.1	0.4	0	0	0
Father's health	Good	85	55.8	57.1	70.9	0
	Fair or poor	6.5	5.4	3.6	3.4	0
	Missing	8.5	38.9	39.3	25.7	100
Father ever in jail	No	98.8	89	78.6	93.9	100

Table 4: Descriptive Statistics for Analytical Variables in the FFCWS Initial Sample by Family Structure at Birth (*continued*)

		Married or Cohabiting two parents	Lone-parent	Complex, step or blended family	LAT	Missing
	Yes	1.2	7.7	14.3	5.4	0
	Missing	0.1	3.3	7.1	0.7	0
Child's gender (%)	Girl	48.4	46	32.1	45.3	0
	Boy	51.6	54	67.9	54.7	100
	Missing	0	0	0	0	0
Low-weight birth (%)	Yes	7.8	11.6	10.7	8.8	0
	No	89.5	86.3	82.1	87.8	100
	Missing	2.7	2.1	7.1	3.4	0
Twins at birth (%)	No	98	98.6	92.9	98.6	100
	Yes	2	1.4	7.1	1.4	0
	Missing	0	0	0	0	0
Alcohol, Tobacco or drug consumption at birth (%)	Used	0.2	0.3	0	0.7	0
	Did not use	99.8	99.7	100	99.3	100
	Missing	0	0	0	0	0

Table 4: Descriptive Statistics for Analytical Variables in the FFCWS Initial Sample by Family Structure at Birth (*continued*)

		Married or Cohabiting two parents	Lone-parent	Complex, step or blended family	LAT	Missing
Living with extended kin or unrelated adults (%)	Nuclear	73	32.4	75	43.2	0
	Extended or complex	27	67.6	25	56.8	100
	Missing	0	0	0	0	0
Mother side siblings in household (%)	No siblings	39.4	31.6	28.6	43.2	0
	Siblings present	60.6	68.4	71.4	56.8	100
	Missing	0	0	0	0	0
Violence by partner against mother (%)	No	68.7	81.4	39.3	66.9	100
	Yes	26.2	18.6	7.1	33.1	0
	Missing	5.2	0	53.6	0	0
Mother does paid work (%)	No	41.2	41.3	53.6	50	0
	Yes	48.7	47.6	39.3	37.2	0
	Missing	10.1	11.1	7.1	12.8	100

Table 4: Descriptive Statistics for Analytical Variables in the FFCWS Initial Sample by Family Structure at Birth (*continued*)

		Married or Cohabiting two parents	Lone-parent	Complex, step or blended family	LAT	Missing
Changed home (%)	Did not moved	43.9	52	42.9	48	0
	Moved	33.1	25	42.9	29.1	100
	Missing	23	23	14.3	23	0
Lives on rent (%)	Own house	36.8	35.1	21.4	33.1	0
	Renting	62.5	64	78.6	64.9	100
	Missing	0.7	0.9	0	2	0
Income to poverty rate categories	0-49%	12.7	27.8	32.1	19.6	100
	50-99%	12.3	20	10.7	19.6	0
	100-199%	24.2	28.2	25	24.3	0
	200-299%	17.4	14.4	10.7	13.5	0
	300%	33.4	9.6	21.4	23	0
	Missing	0	0	0	0	0
Nominal equivalized household income	Median	15910	8839	9856	11599	3750

Table 4: Descriptive Statistics for Analytical Variables in the FFCWS Initial Sample by Family Structure at Birth (*continued*)

		Married or Cohabiting two parents	Lone-parent	Complex, step or blended family	LAT	Missing
Mother'S cognitive skills (PPVT std. score)	Median	93	87	85	85	-

*Note:*  
Own elaboration. Mother's employment is measured one year after birth

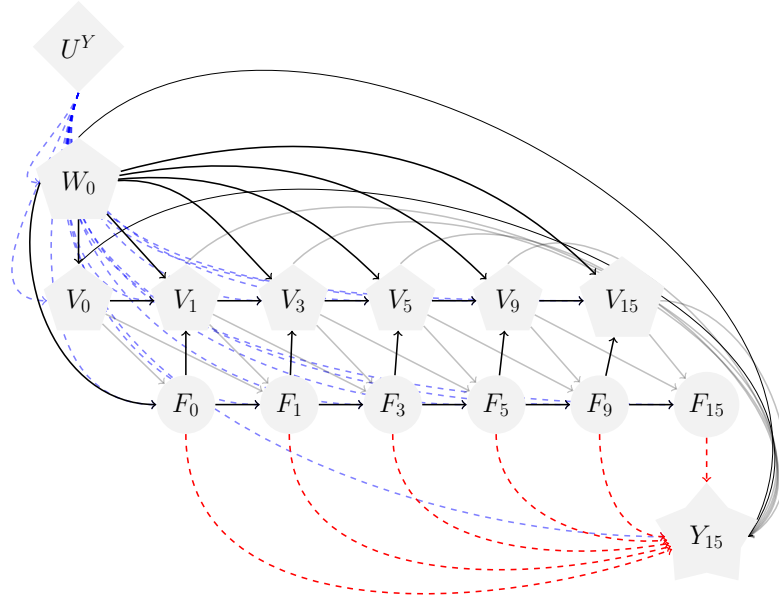


Figure 8: Simplified DAG for the family instability hypothesis and the Selection Hypotheses

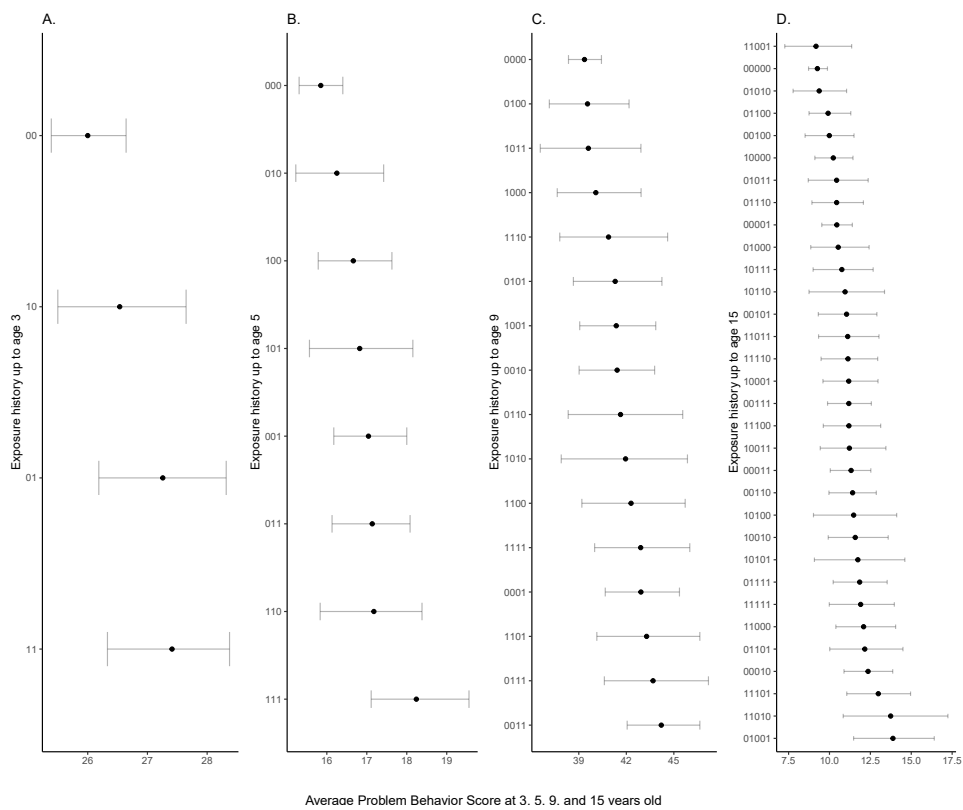


Figure 9: Average problem behavior at various ages following all potential exposures

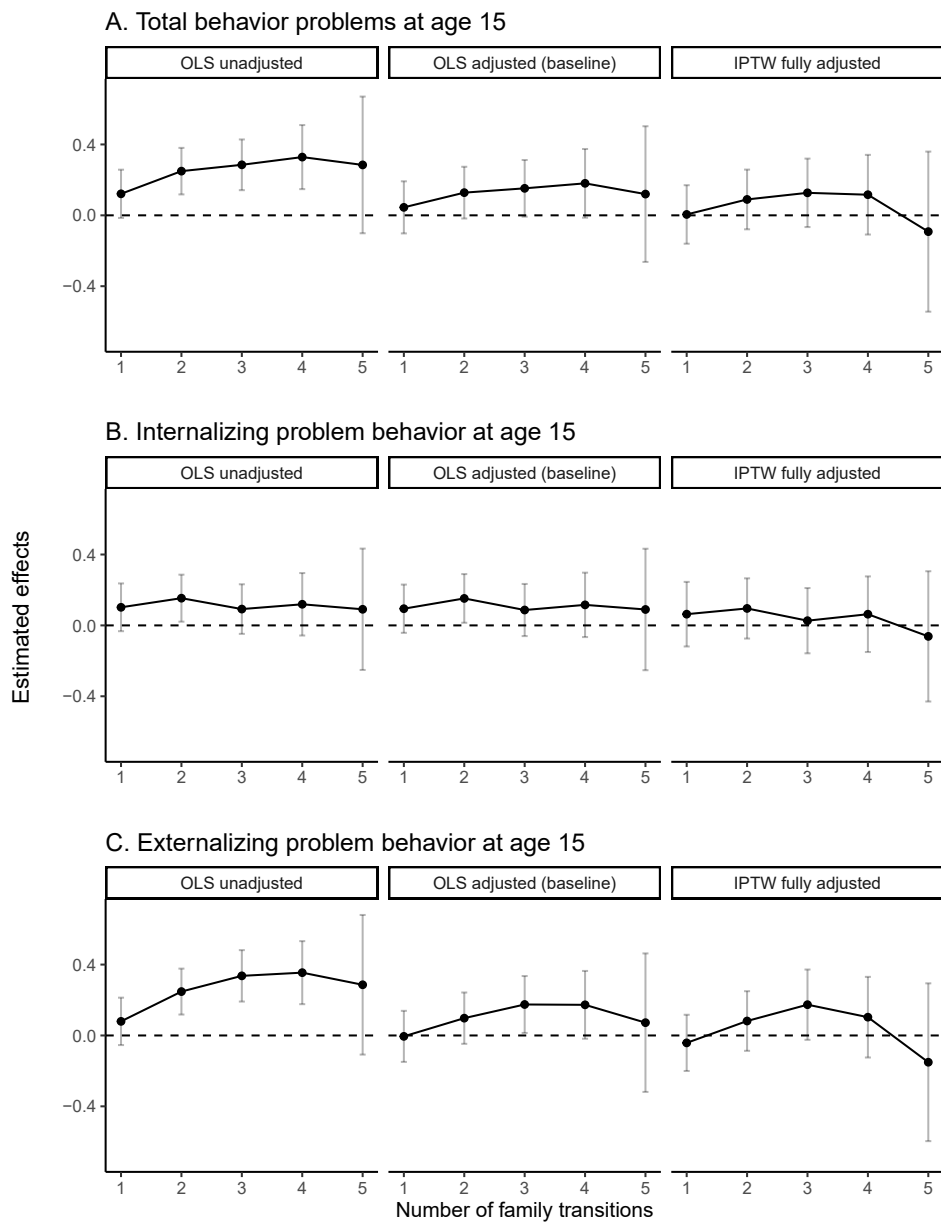


Figure 10: Estimated cumulative effects of repeated family changes on problem behavior at 15 years old

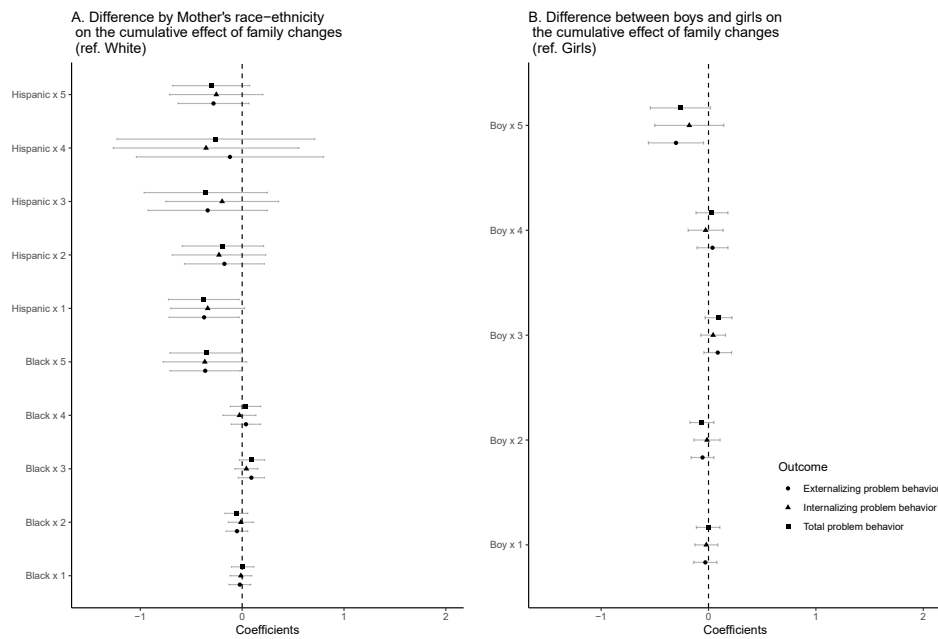


Figure 11: Moderating effect of cumulative changes in family structure on problem behavior at age 15 by child's mother racial-ethnic category and child's gender assigned at birth

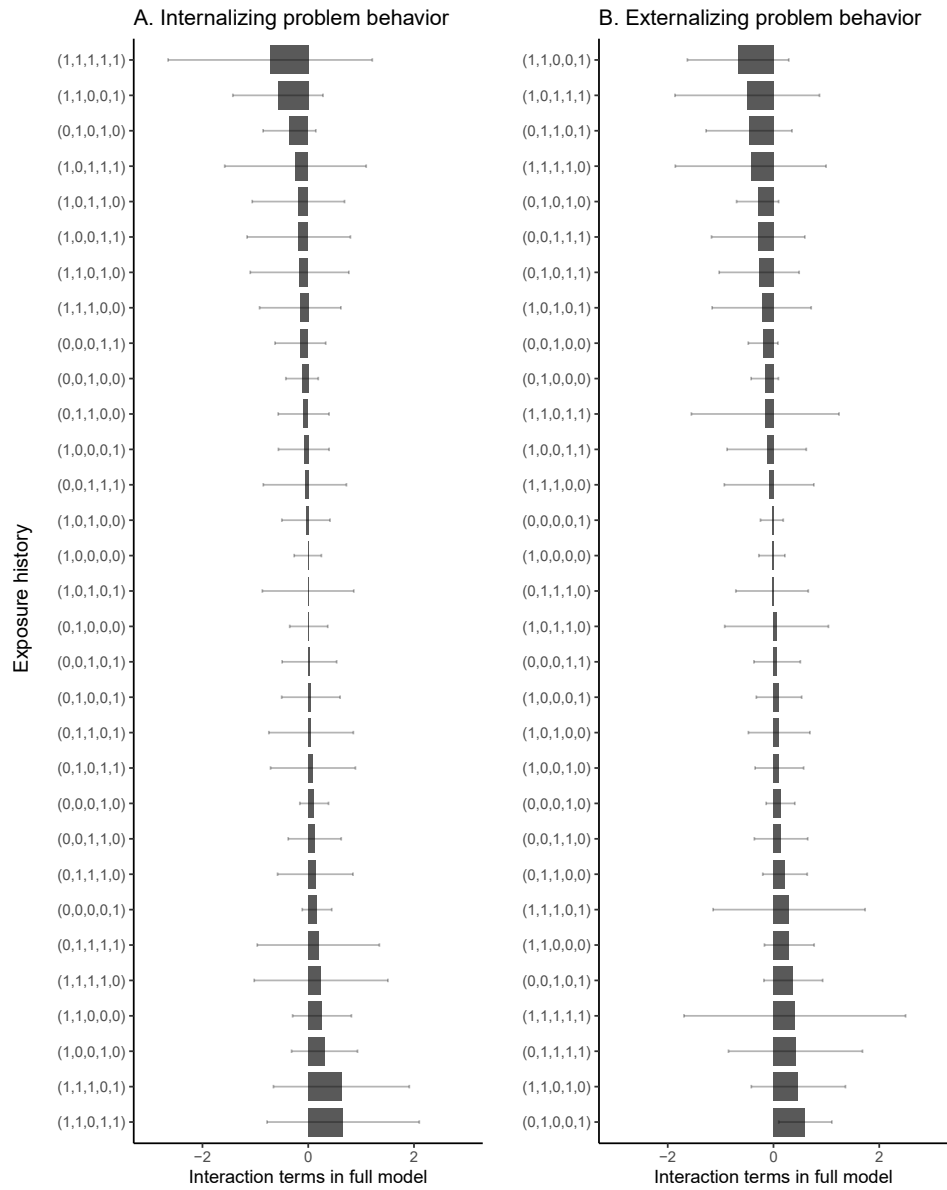


Figure 12: Joint effect family instability exposure trajectories up to 15 years old on the internalizing, externalizing and total problem behavior of adolescents [ref. stable families (0,0,0,0,0)]

## **CHAPTER 4 - What can parents do? The causal mediating role of parenting in explaining SES differences in children's language development**

*This paper has been published as: Rodríguez Sánchez, A. (2023). What can parents do? The causal mediating role of parenting in explaining SES differences in children's language development. Journal of Family Research, 35, 53-84.*

*This is the accepted manuscript. The final version is available at <https://doi.org/10.20377/jfr-688>*

There is a strong association between low parental socioeconomic status (SES) and many adverse outcomes in childhood that carry on into adulthood (Greg J. Duncan et al. 1998). One of the most consistently investigated child development markers affected by low SES is children's language development (Romeo et al. 2018). Research in Germany has shown that SES differences in cognitive development, of which language development is an important part of, emerge well before children enter school and tend to remain stable throughout children's school careers (Skopek and Passaretta 2020). Substantial inequalities in children's developmental trajectories have been observed in other countries as well, but parental SES influences children's wellbeing both directly and indirectly. In the US, for example, where differences in children's language development by parental SES have been long investigated (Greg J. Duncan, ZiolGuest, and Kalil 2010), researchers have begun to narrow in on one salient mechanism: the differences in how low-SES and high-SES parents care for and rear their children (Fomby and Musick 2018).

Concerns over children's wellbeing have taken researchers to closely examine internal family dynamics and in particular caregiver-child interactions

(Mayer et al. 2019), as these might mediate, or indirectly explain, the effects of parental SES. The literature on parenting often distinguishes between three general types of behaviors parents engage in with their children (Spera 2005): parenting styles, parenting practices or daily activities, and parental investments (Doepke, Sorrenti, and Zilibotti 2019; Mayer et al. 2019). These can be considered different dimensions of parenting (Maccoby 1994), though there is a large overlap between them. Parenting style describes the type of parent-child interaction that occurs on a regular basis, and it has been classified into five general types: authoritative, emotional, sensitive, detached, or authoritarian (Baumrind 2005; Kuppens and Ceulemans 2019). Parenting practices, instead, refer to the daily activities done with the child, such as attending to the child's basic needs or promoting the language and reading skills of the child, and their socio-emotional development (Cobb-Clark, Salamanca, and Zhu 2019; Kalil, Ryan, and Chor 2014), as well as the cultivation of literacy at home (Melhuish et al. 2008). Parental investments, although often fused together with parenting styles and practices, refer more to the time and financial resources parents invest in their children in the form of goods/services such as childcare, books for children, or other cognitively stimulating activities (Doepke, Sorrenti, and Zilibotti 2019; Miller et al. 2020).

Parenting covers a large range of practices and activities that parents do with their children, which makes it difficult to exactly differentiate them. Hoff (2013) and Romeo et al. (2018) trace parental SES differences in children's language skills to gaps in some parenting dimensions. For example, the effects of authoritarian or detached parenting styles (T. W. Chan and Koo 2011), as well as the effects of the differences in reading-time with children (Kalil, Ryan, and Corey 2012), and the effects of low or insufficient parental investments (Doepke, Sorrenti, and Zilibotti 2019) do explain to some extent structural inequalities by SES. In fact, such types of

parenting are more characteristic of low-SES families. This is one reason behind the focus of parenting interventions and early childhood programs on teaching low-SES mothers parenting skills deemed necessary for effective child development (Ayoub, Vallotton, and Mastergeorge 2011; James J. Heckman, Humphries, and Kautz 2014; Price 2010). However, and even though specific types of parenting affect children, little is known as to what share of the total effect of SES is mediated through parenting, a quantity of relevance for early childhood public policy.

Whatever that quantity may be, previous studies suggest the link between SES and children’s cognitive development cannot be fully accounted for by parenting (Sullivan, Ketende, and Joshi 2013), which is itself suggestive of alternative pathways. In this study, I explore and estimate the mediated share of the SES effect on language skills that *goes through* parenting, broadly understood (Berendes et al. 2013). To achieve this, I follow the suggestion of previous research and use a broader conception of parenting when estimating these effects, instead of focusing on a single parenting dimension (Fomby and Musick 2018; Mollborn 2016) – an approach that is riddled with identification problems, as I explain in the paper. I estimate the share of the SES effect that is jointly mediated through multiple parenting mediators employing interventional mediation analysis, as explained and developed in T. J. VanderWeele and Tchetgen Tchetgen (2017), to best meet the multiple assumptions on which causal mediation analysis depends on (Nguyen, Schmid, and Stuart 2021).

## **Background**

Various later child wellbeing and development indicators, such as memory skills, literacy, and general school performance, heavily depend on early language development (Shonkoff 2011). Therefore, to narrow social inequality in school performance and academic achievement (T. Schneider

and Linberg 2021), researchers and policy makers need to understand which mechanisms generate early differences in language skills by SES. Studies show parenting is, without a doubt, an active force in children's developmental trajectories (Grusec 2011; Maccoby 2000). Two recent studies estimate the association between SES and language development in the German context (Skopek and Passaretta 2020; Attig and Weinert 2020). These studies employed the same data used in this paper and find a long lasting impact of SES – parents' education – on language, and on cognitive development more generally, as well as strong differences in parenting practices or joint activities by SES.

Parenting is thought to constitute one of the most salient processes through which social inequalities are reproduced from one generation to the next (Guo and Harris 2000). Among these processes, family influences like parenting are but one important social determinant of children's development (Maggi et al. 2010). Although the overview of Hoff and Laursen (2019) discuss parenting as a moderating factor in the association between SES and children's language development, it might be more appropriate to consider parenting as a general mediating mechanism explaining SES-related differences in child development (Weininger and Lareau 2009). Both the family stress model and the family investment model suggest parenting is, however, if not the main, certainly one important mechanism linking children's development trajectories and SES (Conger, Conger, and Martin 2010; Greg J. Duncan et al. 1998). Although parenting dimensions are conceptualized as different though related aspects in the rearing of children, recent research suggests that these are not independent of one another, and that, therefore, their effects on children's development should perhaps be considered jointly (Elliott and Bowen 2018; Mayer et al. 2019).

## **Parenting and the development of language skills**

In addition to the SES and parenting association, there are several pathways linking parenting and various aspects of language development in early childhood (Pace et al. 2017). The effect of various parenting dimensions on the cognitive development of children has been recently summarized by Ulferts (2020), including multiple studies and meta-analyses. A recent meta-analysis of studies evaluating the linkages between parenting and language development suggests that sensitive parenting facilitates language and learning in children in a myriad of ways (Madigan et al. 2019). In particular, evidence from randomized controlled trials show that the effect of interventions on parental sensitive responsiveness on children's language in economically disadvantaged families is stronger than what is found in high-SES families (Madigan et al. 2019). Moreover, S. C. Perkins, Finegood, and Swain (2013)'s review of basic brain functions and areas related to language development further suggest that these are directly and indirectly affected by exposure to low SES contexts and parenting (Merz et al. 2019; Kuhl 2004).

In this respect, word variety, a higher complexity of language, as well as more frequently asking children questions eliciting a response from them, which are more characteristic of high SES parents and are linked to specific parenting behaviors (Vernon-Feagans et al. 2012; Ochs and Kremer-Sadlik 2015), would be more beneficial for early language skills. Early word learning, for example, is often the outcome of children's engagement in joint activities with their caregivers (Hart and Risley 1995; Kuchirko 2019), an effect that is enhanced when caregivers are responsive to children's needs and solicitations (Hoff 2003). There is evidence that low educated mothers display a less responsive type of communication with their children, which suggests this may be one explanation of the lower verbal abilities found

among low-SES children (Dollaghan et al. 1999). Similar findings have been observed in Germany. Attig and Weinert (2020) show that various characteristics of the home learning environment of children vary by SES, with negative impacts on children's language skills.

Other research suggests the effects of parenting might be heterogeneous as a consequence of families being embedded within specific contexts around which family life is organized (Manstead 2018). The effects of certain parenting styles may depend on broader cultural and local contexts in which parenting is embedded. The effects of the same parenting practice of, for example, reading a book to a child, might depend on the expectations that peer-groups or families have for a child, and a myriad of additional effects that might enhance, reinforce or suppress, and therefore discourage, the effects of parenting styles, practices, or investments in those contexts.

Regarding specific parenting dimensions, various pioneering works on the child development, parenting literature suggest strong interactions between these different dimensions, which are thought to generate differences in language development. Authoritative styles might be more important in specific environments, particularly those of low SES families (Sorkhabi and Mandara 2013). Parenting practices' effectiveness in promoting language development, on the other hand, may depend on the type of parenting style (Darling and Steinberg 1993). Moreover, parenting styles could be an underlying cause of specific parenting practices and parental investments, causing parents to engage in more activities with their child and invest more time and resources in them (Doepke, Sorrenti, and Zilibotti 2019).

### **Differences in parenting behavior by SES**

There are multiple mechanisms linking SES and parenting, which explains why SES differences in parenting could be an important reason behind

language skill gaps. Most of these mechanisms operate through the stress associated with raising a child in low-resource contexts (Bradley and Corwyn 2002). Economic insecurities associated with low SES, added to the diffusion of the ideology of intensive parenting, are major sources of stress affecting parental wellbeing and parenting (Nomaguchi and Milkie 2020). Moreover, Doepke and Zilibotti (2017) show important differences in parenting by the overall socioeconomic status of the country, and substantial differences in time involved in parenting by parents' educational level. Hoff and Laursen (2019) conclude that evidence suggest multiple causal pathways are at work between these two factors (Merz et al. 2020). The variables involved in these pathways - which include external factors, such as circumstances in which parents live, and internal ones, such as parents' socio-demographic characteristics - act jointly to generate differences in parenting (Hoff and Laursen 2019).

The main theoretical perspectives explaining the connection between SES, parenting, and children's cognitive development rely on a set of complex interactions between multiple parenting mechanisms (Hoff and Laursen 2019). The family investment model, for instance, considers that differences in resources and parental investments explain differences in children's cognitive development (Greg J. Duncan et al. 1998). The family stress model, instead, centers on the stress that is caused by low resources, and on how this form of stress negatively affects parents' behavior when rearing and educating their children (Conger, Conger, and Martin 2010). Moreover, research on the parenting package (Fomby and Musick 2018), as well as on the effects of developmental ecologies (Mollborn 2016), suggest the mediating linkages between SES and parenting might be far more complex than currently understood. For Bronfenbrenner and Morris (2007)'s bio-ecological model, parenting is a proximal process within the family, which influences children's development through multiple pathways.

However, the effects of these pathways are embedded in networks of peers, neighborhoods, and larger institutions, thus hypothesizing a much less mechanistic role for parenting in child development. For example, low-SES parents might have a heavier workload and receive lower wages, which implies higher levels of tiredness, stress, and, perhaps, less leisure time to spend with their children, all of which may produce a less effective type of parenting (Milkie, Raley, and Bianchi 2009). In contrast, high-SES parents are able to hire or pay for extra parenting provided by others in the form of cognitively stimulating activities with care professionals, even from early on (Schober and Schmitt 2017); which may improve the parent-child relationship by taking away some of the burden of care work from parents' shoulders.

Parenting is something done by unequal parents and in unequal circumstances (Lahire 2019). For example, studies have shown that middle-class parents engage in parenting practices that have been characterized as *concerted cultivation*, positively reinforcing middle-class children's skills, whereas working-class and working-poor-class families' parenting is more similar to the *accomplishment of natural growth*, a different kind of parent-child interaction that may not provide these same benefits (Lareau 2011: pp. 238-239). On top of that, other qualitative research suggests that social inequalities in children's language skills result from the interaction between social class origins and teachers' or caregivers' differential expectations about children's behavior and language (Millet and Croizet 2016). Parents are not only unequally endowed with resources and skills for child rearing, they are unequally constrained by factors such as time and working conditions that also have effects on their parenting (Bradley and Corwyn 2002). Given the highly complex interactions between parenting mediators, and the many confounding factors affecting them, there is reason to doubt parenting could mediate close to the full share of

the SES total effect as theoretically expected (Milkie, Nomaguchi, and Denny 2015; Sullivan, Ketende, and Joshi 2013), in fact, it is more likely that this share is rather small.

### **SES as a causal factor in parenting and children's language development**

Most studies in the parenting literature reduce parenting to one of the child's parents educational attainment (Kalil, Ryan, and Corey 2012), often that of the mother's. This reductionist approach, although encouraged by some highly popular research (Hoff and Laursen 2019; Cunha, Heckman, and Schennach 2010; Goldthorpe 2010), obscures rather than clarifies many of the ways in which parenting actually operates as a proximal process. The reductionist approach is not universally accepted by all research on parenting (Thaning and Hällsten 2020; Savage et al. 2013; Sullivan, Ketende, and Joshi 2013; Goldfeld et al. 2018). Other approaches that consider additional SES related variables, such as income, wealth, occupational attainment, welfare dependency, etc., and combine them, have also received attention in the parenting literature (Ishizuka 2019), and in explaining gaps by SES in other child wellbeing measures (Berge et al. 2010).

In socio-linguistic research, it was early recognized that families use of language is fundamentally a function of their SES or social class (Bernstein 1964; Bourdieu 1991). Children from different SES backgrounds acquire different communicative abilities, but these are not a simple additive function of SES as assumed by the reductionist approach. I argue that it is necessary to include both parents' resources in understanding the effects of family SES, as it has been recently discussed when studying siblings correlations on educational outcomes (Thaning and Hällsten 2020).

There are important disadvantages in the reductionist approach vis-à-vis its

alternative – that of combining the different dimensions (Hauser and Warren 1997; Goldthorpe 2021; Hollingshead et al. 1975) – which I address in this paper. Mother’s educational attainment correlates strongly with other SES dimensions – and with those of the father – and it has been found to be the most predictive of SES factors for children’s language (Kalil, Ryan, and Corey 2012). However, that approach does not correspond with the many nuances reviewed above of parenting as a function of SES. For example, occupation, income, and wealth, are fundamentally functions of educational attainment, which would make it a type of mediator that also affects parenting. These factors may also moderate the effects of education, with differences in their impact. There is little research on the effect of socio-occupational class, but household income has been shown to be of little relevance (Berger, Paxson, and Waldfogel 2009), and recent studies suggest wealth may have an important independent, additive effect (Dräger 2022). However, when children are born, they experience all the dimensions of their mother and father’s SES simultaneously, not just one dimension at a time, so it is difficult to try to separate them (Bihagen and Lambert 2018). Each SES dimension affects what parents know or are able to do with their children (e.g., networks, learning from peers, etc.) and, therefore, may affect the functioning of the educational attainment dimension, even in systematic ways (Heinrich 2014; D. Schneider, Hastings, and LaBriola 2018). Moreover, given that not one single dimension can operate equally, the combined approach has the advantage of acknowledging important differences caused by families’ divergent class backgrounds (Gillies 2005; Lahire 2019), in the direction of the individuation of social class.

Notwithstanding these nuances, the major disadvantage of the reductionist perspective is that employing mother’s educational level to approximate SES ignores potential measurement error in this variable. Educational attainment of the mother does not fully capture the construct of SES and it

does so with systematic error, for example, by ignoring father's educational level or assuming that different educational degrees, such as doctor versus school teacher, at the same educational level, are entirely comparable. For example, the occupation of both parents may be relevant for language development when parents work in fields such as education or health versus manual occupations with little involvement of verbal abilities, beyond parents' educational levels. Equally so, parents with high financial resources, but low educational attainment, may compensate for lacking educational credentials through private tutoring, by engaging in cultural activities themselves, or by enrolling their children into verbally intensive activities or social networks. Therefore, and although the SES-related individual variables may be used in place of a latent construct, these do not fully represent, nor override, the construct of SES.

One SES dimension does not replace another one and cannot solely stand for the other ones either (Thaning and Hällsten 2020). The literature on parenting so far offers no answers to these critical issues because underlying the suggestion of using only mother's education there is no causal model being hypothesized, or only one following a conception of education as human capital (G. S. Becker 2009), supposedly uniform and exchangeable. In contrast to the reductionist approach, and considering that SES is an unobservable and theoretical construct (Lahire 2019; Savage et al. 2013; Goldthorpe 2021), an inductive or formative methodology seems more appropriate. The debates around the measurement of SES do suggest that a measurement model could incorporate all the dimensions and address the limitations of the reductionist approach – of which no advantage has been shown (Thaning and Hällsten 2020).

## **Data and Methods**

### **Causal interventional direct and indirect effects**

In previous research, the estimation of direct and indirect effects of specific parenting dimensions has been based on so-called “third variable analyses” (Attig and Weinert 2020; T. Linberg et al. 2019), also known as statistical mediation or statistical decomposition methods – in contrast to causal mediation analysis – such as structural equation models or Oaxaca-Blinder decomposition methods (Nguyen, Schmid, and Stuart 2021; Guo and Harris 2000). However, recent advances in causal mediation analysis stress that the identification of mediating pathways relies upon various assumptions that often go untested in empirical studies (T. VanderWeele 2015), and that are unrealistic in almost all empirical applications. Statistical decomposition methods provide descriptive analyses of the exposure-mediator-outcome associations, but these do not warrant a causal interpretation. Furthermore, these methods only consider the case of a single exposure, that is one mediator at a time, and assume independence among all possible individual mediating mechanisms, which is the reason why the sum of all “indirect” effects obtained with such methods can sometimes exceed the total effect under investigation. The reason for this is that statistical decomposition analyses do not explicitly address the problems of confounding, which bias the magnitude of the mediating pathways, and which should be addressed by estimating counterfactual mediators and outcomes.

Parenting as a mediating mechanism, as discussed in the background section, is hypothesized to be a mediating process involving multiple, distinct mediators that interact with each other, and which follow multiple pathways. Therefore, studies of parenting require a more encompassing exploration of the role of parenting, one that considers the various dependencies and pathways involved in the SES, parenting, and language associations. As an

example, Figure 13 displays many of the complex hypothesized interactions that I have discussed in the Background section. The different parenting mediators (here simply denoted as parenting practices  $M^P$ , parenting styles  $M^S$ , and parental investments  $M^I$ ), interact with one another in ways that are largely unknown, and probably context dependent. Although these parenting dimensions can be thought of as mediators, studies suggest strong interactions with one another, something which should be considered in estimations of the mediated share. Not only that, but parenting is itself a function of multiple factors beyond SES, which should be considered simultaneously when estimating mediation. These are factors which are affected by SES, the exposure of interest, but which also have direct and indirect effects on the language development indicators,  $Y$ . There are two major sources of this type of confounding: a general confounding mechanism  $C$ , such as parents' ages, and an exposure-induced confounding mechanism  $C^M$ , such as smoking during pregnancy, low birthweight, parental working hours or use of child care arrangements. The existence of exposure-induced confounders has been neglected by most previous research, yet it is especially problematic given that these confounders are affected by SES, affecting equally all parenting mediators and children's language development.

In contrast to statistical decomposition methods, causal mediation analysis addresses these issues, and emphasizes the problems of identification. Four main assumptions are necessary for providing a causal interpretation of mediation. The first assumption is that there is no unmeasured confounding of the exposure-outcome association  $Z \rightarrow Y$ , where  $Z$  stands for SES and  $Y$  for children's language skills. In mathematical notation, this assumption can be written as  $Y(z, m) \perp\!\!\!\perp Z|C, \forall z, m$  with  $z$  being the observed level of SES and  $m$  the values taken by the mediators (e.g., a family SES is affected by the child's grandparents' SES which also affects their verbal skills). Second, we need to assume there is no unmeasured confounding of the

mediator-outcome association,  $M \rightarrow Y$ , or that  $M(z) \perp\!\!\!\perp Z|C, \forall z$ , where  $M(z)$  is the counterfactual value of the mediators – the parenting styles, practices and parental investments – had the exposure been set to  $z$  possibly contrary to fact (i.e., a highly sensitive mother that attends to her child’s needs although from a low-SES background). The third assumption is that no unmeasured confounding factors exists between the exposure-mediator association,  $Z \rightarrow M$ , or that  $Y(z, m) \perp\!\!\!\perp M|Z = z, C, \forall z, z^*, m$ , with  $z^*$  being the counterfactual level of exposure SES different to one actually observed  $z$  (e.g., problem behavior in children may affect parenting and also language development) Finally, the fourth assumption states that there are no confounders of the mediator-outcome association,  $M \rightarrow Y$ , that are affected by the exposure, that is, that they are also caused by  $Z$ :  $Y(z, m) \perp\!\!\!\perp M|Z = z, C, \forall z, z^*, m$  (i.e., smoking or drinking during pregnancy is associated with low-SES families, it may affect the type of parenting needed, and it also affects children’s development).

Once explicitly stated, it becomes clear that meeting any one of these assumptions in the context of parenting as a mediating mechanism between SES and children’s language skills is rather difficult. As discussed above, the fourth of these assumptions is particularly problematic. We can be fairly sure that the no exposure-induced mediator-outcome confounding assumption does not hold for the mediating mechanisms we have in mind. Here two examples: a) Smoking during pregnancy or a premature birth are affected by maternal education, but both smoking and having had a premature child may affect the type of parenting done by caregivers and the children’s language or cognitive development; b) Dual-earner families, in which both parents work full-time, may have less time to engage in highly intensive parenting practices, but full-time work, especially in high paying occupations, may also allow families to purchase time-intensive activities for their children or higher quality childcare. Moreover, all major surveyed

theories in the parenting literature – the family stress model (Conger, Conger, and Martin 2010), the family investment model (Cobb-Clark, Salamanca, and Zhu 2019), and the bio-ecological model (Bronfenbrenner and Morris 2007) – describe many more potential mediating pathways.

If SES affects a parenting dimension, say parental investments, that in turn affects another one, such as parenting practices, then the fourth assumption required for causal mediation does not hold for these single mediating pathways. For example, research has shown sensitive mothers perform more time-intensive parenting practices with their children, and that high-SES mothers adjust their parenting to fit the child’s needs (A. Linberg 2018; Kalil, Ryan, and Corey 2012). Children from low-SES backgrounds may spend more time with their mothers - who are more likely to be unemployed, out of the labor force, or working part-time, but the time-intensive parenting with low-educated mothers may not be as effective in the teaching of language skills if the mother-child interaction does not correspond to a sensitive parenting style. Also, parental investments in the form of books or cognitively stimulating toys may affect how often children are exposed to literary content and stimulating activities (Greg J. Duncan et al. 1998), and through such pathways to more or less frequent reading or less time with sensitive mothers. Therefore, parenting is akin to a causal sequential chain that lacks a clear specification because it is unclear what is causing what amid the different parenting mediating mechanisms. However, the identification of individual pathways is troublesome and, following causal mediation analysis, this sequential chain may not be empirically testable.

An estimation of the mediated share of the SES effect that goes through parenting requires a particular kind of causal mediation analysis that accounts for exposure induced mediator-outcome confounding and mediator-mediator effects (T. VanderWeele and Vansteelandt 2014; T. J. VanderWeele and Tchetgen Tchetgen 2017). An alternative approach to the

mediating effects of parenting – broadly understood – is to take on a so-called “interventional perspective” that considers a joint version of the individual mediating pathways (Nguyen, Schmid, and Stuart 2021; T. J. VanderWeele and Tchetgen Tchetgen 2017). We let  $\bar{M} = (M^I, M^P, M^S)$  be a vector containing all the different mediators of parental investment, joint practices, and parenting style, and define  $G^{\bar{M}}$  as the conditional distribution of the mediators vectors, which is conditional on the common factors that confound the association between *SES* and children’s language skills. What would be the effect of intervening or modifying the overall distribution of the mediators,  $G^{\bar{M}}$ ? The interventional effect is of relevance for understanding ideal interventions on parenting, such as the interventions proposed in the work of Mayer et al. (2019), which are aimed at reducing the differences between the parenting of low-SES mothers. One plausible research question that practitioners and researchers would be equally interested in is: what *would* happen to the average low-SES children language skills if we were to replace the parenting experiences these children normally obtain in their family for a randomly selected experience from the distribution of parenting vector  $G^{\bar{M}}$  observed among high-SES parents? An answer to such a counterfactual type of question – involving a sustained intervention on parenting – could tell us what benefits we could expect from parenting interventions that equalize the parenting done in families of different SES, especially as these are being considered or implemented in Germany as well (Walper 2021; Cina et al. 2006).

In causal mediation notation, and as fully explained and developed in T. J. VanderWeele and Tchetgen Tchetgen (2017), the total effect (*TE*) of the exposure of SES can be decomposed into a Pure Natural Direct Effect  $PNDE = E[Y(z, \bar{M}(z^*))] - E[Y(z^*, \bar{M}(z))]$  and a Total Natural Indirect Effect  $TNIE = E[Y(z, \bar{M}(z))] - E[Y(z, \bar{M}(z^*))]$ . Therefore,  $TE = PNDE + TNIE$ . Here  $z$  denotes a value taken by the exposure  $Z$ ,

which in this case would correspond to a given SES level, and  $\bar{M}(z)$  is the value the mediators take for that given SES level (i.e., the values taken along the parenting pathway for a low-SES parent, such as insensitive parenting, few joint activities, and low parental investments). The counterfactual quantity  $Y(z, M(z))$  can be interpreted as the child's language score after growing up in a given SES context  $z$  (e.g., very low SES), had the child have had the mediators  $\bar{M}$  taken the actual value  $\bar{m}$  corresponding to that context. However, these are not identifiable quantities as explained above (Nguyen, Schmid, and Stuart 2021). Other approaches, such as 'en bloc' mediation (T. VanderWeele and Vansteelandt 2014), have the advantage of identifying natural direct and indirect effects, but they do not allow us to separate the part of the total effect that goes through other mediating mechanisms not involving parenting (i.e., the exposure-induced confounders of the mediator-outcome association), and instead mix the mediated share going through parenting with the mediated share going through these other pathways.

Instead, the interventional or randomized counterparts of these causal mediation parameters can be estimated despite the presence of exposure-induced confounders of the mediator-outcome association, and, importantly, in the presence of more than one single mediator, as in  $\bar{M}$ . The interventional analogue of these parameters is based on the idea of randomly drawing the values of the mediators from the counterfactual distribution – that of the comparison group, and are defined as follows. The  $rPNDE = E[Y(z, G_{z*}^{\bar{M}})] - E[Y(z*, G_{z*}^{\bar{M}})]$  is the effect of the exposure that does not involve the mediators (the randomized/interventional analogue of  $PNDE$ ); whereas  $rTNIE = E[Y(z, G_{z*}^{\bar{M}})] - E[Y(z, G_{z*}^{\bar{M}})]$ , the effect involving a change in the mediators, is the randomized/interventional version of the Total Natural Indirect Effect. The sum of these two quantities provides again the Total Effect, which makes sense for a decomposition.

With the decomposition into a direct and indirect effect, one can then obtain the proportion mediated by this randomized/interventional quantities, defined as  $rPM = \frac{rTNIE}{TE}$ .

The counterfactual distributions, from which  $rPNDE$  and  $rTNIE$  are computed, are obtained through the imputation approach. These imputations, in turn, are based on a set of statistical models fitted to each variable in the mediating set, in this case, each parenting indicator, regressed on the exposure, the exposure-induced confounders, and other confounders, as well as a model for the outcome (children’s language skills), and, importantly, a model for the exposure-induced confounders. Predictions from these models are used to compute the g-formula (Robins 1986). Standard errors and confidence intervals are obtained by means of bootstrapped results and its percentiles, respectively.

One drawback of the randomized/interventional approach, as well as of the traditional statistical decomposition methods, is that identifying individual pathways is no longer possible without making strong assumptions – unlikely to hold for the case of parenting. However, this is currently the best we can possibly do when trying to decompose a total effect into its direct and indirect pathways (Nguyen, Schmid, and Stuart 2021).

Notwithstanding, an advantage of interventional mediation over joint mediation analysis and, especially, over statistical decomposition methods is that we are able to exclude the share of the effect of SES that operates through observed confounders of the mediator-outcome association affected by exposure (e.g., working hours, childcare arrangements, behavior problems, premature birth, smoking during pregnancy, etc.), which do not correspond to the effect of interest. To estimate the  $rPNDE$ ,  $rTNIE$  and the  $rPM$  I employ the mediational g-formula proposed in T. J. VanderWeele and Tchetgen Tchetgen (2017), and as implemented in the R package *CMAverse* (Shi et al. 2021). This method uses the imputation

approach with 200 bootstraps samples for estimating standard errors of the interventional/randomized quantities.

***Missing data and panel attrition*** Missing data and systematic panel attrition may generate bias in the estimates of interest in this paper. To deal with these two issues, multiple imputation by chained equations with 56 imputed data sets was performed (Young and Johnson 2015). The method CART was used to impute missing values because it has been shown to ensure the best possible predictions within the observed value range of each variable (Burgette and Reiter 2010). All analytical variables were used in the multiple imputation procedure. After converting the data into wide format, all variables observed at time point  $t - 1$  served as predictors for the missing values at  $t$ , and subsequently those from  $t - 1$  and  $t$  were used as predictors for the missing values at  $t + 1$ , excluding observations lost to follow-up did not affect the findings. No noticeable convergence issues were observed in the resulting multiply imputed values.

## ***Data***

This study uses data from the National Education Panel Study (doi:10.5157/NEPS:SC1:6.0.0), a random sample of Germany-born children, a cohort of newborns in Germany that has been followed for over seven consecutive years. The data sampling was based on official register data on births that occurred between 2011 and 2012 in Germany (NEPS-SC1, Blossfeld, Roßbach, and Maurice 2011). NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LifBi) at the University of Bamberg in cooperation with a nationwide network. The sample was generated using a

complex random study design and in each follow-up survey, children's characteristics, and performance in standardized tests were recorded along with parents' characteristics and their parenting. The analytical sample was established as follows. From the original sample of NEPS-SC1  $N = 3418$ , I excluded all children with a migration background (i.e., being second or third generation,  $N = 1589$ ) to arrive at a final sample of  $n = 1892$  children for which both parents have no migration background.

Although this exclusion prevents me from exploring the topic of parenting among the Germany-born children of migrants, standardized language assessments, such as the ones used in this study, have been shown to be biased against children from minority-language homes (Leśniewska, Pichette, and Béland 2018). The language assessments used in this study are based on the German language, the inclusion of children with a migration background from a minority-language home may understate the effects of parenting, as understood in this paper, because for this population group parenting may not appear to readily translate into language development as measured in German (i.e., although, if language skills were measured in the minority-language of the home the child grows up, the effects might be comparable). Therefore, in order not to confuse the minority-language background with low-skills in language, this exclusion is necessary.

### **Exposure: Latent class analysis to measure SES**

One can think of approaches to the measurement of SES such as Latent Class Analysis (LCA, Berge et al. 2010), in which both parents characteristics can be combined, as a form of measurement model of the construct of parental education, or, alternatively, as more appropriately capturing such highly complex and nonlinear interactions among the different factors making up SES. I employ LCA to obtain a less error-prone measure of the family SES (Hagenaars and McCutcheon 2002). LCA is

convenient because we can combine multiple dimensions of SES that are interrelated in complex ways (Conger, Conger, and Martin 2010), and because I can employ the information on both the father and mother. The underlying assumption in LCA is that the individual SES indicators – mothers and fathers’ educational attainment, socio-occupational class, income and wealth levels, and welfare recipient status – are assumed to be conditionally independent given the hypothesized latent class. In this sense, each latent class, a categorical variable, would be what makes those indicators correlate with one another, a correlation that could be interpreted as SES or as a refined measure of children’s parents’ educational background. If the social position of a family is what creates an association between markers of SES – as theoretically developed in the perspective of Bourdieu (1984), then it seems plausible to argue and interpret the latent classes as the socioeconomic position occupied by parents, the unobservable latent SES. The number of latent classes was chosen inductively, according to statistical fit criteria using the Bayesian Information Criteria (BIC), which gave four latent classes providing the best statistical fit (Hagenaars and McCutcheon 2002).

The following parental socioeconomic indicators were used in this analysis: the educational levels of both the child’s mother and father; the occupations of the child’s mother and father following the occupational class structure of Eriksson and Goldthorpe (see Evans 1992); a categorized version of the monthly household adjusted income level by the quartiles of the household income distribution in the sample; a categorized version of the level of net-wealth in the household by quartiles of the distribution as well; and, finally, an indicator of whether mother or father received welfare benefits. Although I interpret these latent classes as the “position” each family occupies in the social structure, another interpretation of them is as more “refined” measures of an underlying construct for which maternal education

stands for (e.g., human capital, cultural capital, etc.).

***Outcome: language skills***

Two standardized language assessments for children are used to quantify language development. First, the Parent's Questionnaire for the Early Diagnostic of Children at Risk 2 (*Elternfragebogen für die Früherkennung von Risikokindern 2*; ELFRA-2) was given to the child's parents when children were between 25-27 months of age. This is a standardized questionnaire containing a list of words that a child might be able to utter by a given age – a list of words that is highly sensitive to specific socio-cultural contexts. The scale has been shown to have adequate reliability values according to the standards but is sensitive to the language spoken at home (Sachse and Von Suchodoletz 2007). This questionnaire is filled in by the child's main caregiver and consists of three sub-scales that assess productive vocabulary, syntax, and grammar abilities in German. The ELFRA-2 is used as a screening test for diagnosing delays in language development in German children aged 24 months (see for an overview of the test Grimm and Doil 2006), but it may serve as a first indirect assessment of language skills in early childhood. I make use of the sub-scale on productive language (**ELFRA-2P**), which is only focused on vocabulary knowledge. Although the ELFRA-2 may be affected by response desirability bias, the scale may still serve as a rough measure of the child's vocabulary size for such a young age (e.g., it resembles and has a high correlation with other more standard language assessments that are only doable on older children).

The second assessment was the Peabody Picture Vocabulary Test Fourth Edition (**PPVT-4**), which was applied at waves four and six when children were between 37-39 months of age and again when they reached five years of age, respectively. The PPVT-4 assesses receptive vocabulary or verbal/language skills in children and adolescents in the German

norm-referenced sample (Roßbach, Tietze, and Weinert 2005). The test contains 228 items divided into 19 sets, each containing 12 items. These items were presented on a tablet displaying four different images, only one of which corresponded to the word. During the test, the child listens to each word and must then select the image corresponding to the word heard. The total number of correct responses is often used to establish the stage of children's language development, but the test is also used as a general measure of cognitive development.

### ***Mediators: Parenting***

Given that parenting involves multiple factors and processes, this study tries to encompass the broadest possible available set of things that parents can do with their children. There are three main parenting dimensions I explore in this paper, these are described in Table 5, which documents at greater detail the construction of each indicator, the rationale behind it, and the parenting dimension each indicator corresponds to.

The approximation to the construct of *parenting style* was based on the ratings given to a recorded and standardized mother-child play situation interaction (A. Linberg et al. 2019; Sommer and Mann 2015). This interaction was videotaped and coded by independent raters according to how characteristic or uncharacteristic certain statements about the interaction apply to the series seen in the video. These statements concern the sensitivity to distress and non-distress present in the child, intrusiveness and detachment from the part of the mother, as well as, stimulation, positive regard, and emotionality. Given that this is an interaction situation, within this paradigm, both mother and child's behavior were rated because multiple reciprocal interactions might take place (A. Linberg 2018). However, only the mother's behaviors were used to approach parenting style. Different strategies have been proposed to classify parenting styles, and

although I explored whether these indicators could fit the theoretical expected typology of mothering (Spera 2005; Darling and Steinberg 1993; Baumrind 2005), the results were not satisfactory, with the large majority of mothers falling in an “average” category. For this reason, I included each of these indicators as a binary variable, re-coding individual items.

*Parental investments* were captured through three different types of questions. First, I employed parents’ reports of spending on activities such as participation in toddler or play groups, baby swimming, music groups, or parent-child programs. Second, I considered that another form of parental investment could be to borrow time from work to spend with the child, as in parental leave. I included a binary indicator for whether the mother and father took parental leave during the first two years after the child’s birth. Third, yet another form of parental investment corresponds to parents buying or organizing childcare by other means, for example, through grandparents, childcare, nanny, au-pair, friends, etc. Therefore, a binary indicator of whether the child spends some time in any of these childcare arrangements was included.

For the dimension of *parenting practices* I make use of an extensive list of joint activities that are performed with the child. These are based on the list of items detailed in Table 7 and refer to the frequency of joint activities with the child. Among these items, I included all information on activities such as reading to the child, dealing with literacy and numeracy activities, such as recognition of words, playing with the child in cognitively stimulating activities, teaching rhymes or songs, etc. These activities could have been performed by the parents themselves or by any other person in the household providing childcare for the child. This simple index was calculated for each wave. Although these are simple additive and standardized scores, they may serve as rough approximations to the frequency with which early literacy activities take place and do not affect

the overall results of the mediation analysis (e.g., the mediation analysis including each individual item, as its own mediator, did not differ from the ones here presented). Related to this dimension, I further included another set of indicators of parents' *cultural capital* activities which included things like buying books, visiting museums, going to the theater, and other cultural activities (all described in Table 7).

### **Exposure-induced confounders of the mediator-outcome association**

One important advantage of the interventional mediation analyses – with respect to statistical mediation or joint mediation – is that we can separate the effects of the parenting dimensions from the factors that simultaneously affect those dimensions and also children's language abilities (i.e., the confounders of the mediator-outcome association affected by SES).

Exposure-induced confounders considered here are factors like whether the child was premature birth, whether the mother smoke tobacco or drank alcohol during pregnancy, the number of months she breastfed the child, whether child born with a low birth-weight ( $< 2500g.$ ), whether the mother experienced postpartum depression, the working hours of mother and father, and, importantly, children's behavioral problems, measured by the Strength and Difficulties Questionnaire (SDQ; Goodman et al. (2000)). These are all factors affected by SES that may further affect children's language skills.

For each factor in this list – which is far from exhaustive, but limited by what is captured by the NEPS SC1 study. However, this decomposition of effects is not as straightforward when we consider the intermediate effects of parenting. Here is important to make a clarification regarding intermediate outcomes of children's language, and the exposure-induced confounders. If early language skills are affected by SES and parenting, and if they have an independent direct effect of later language skills ( $Y_t \rightarrow Y_{t+1}$ ), then early

language becomes an exposure-induced confounder when examining the mediating effects of parenting. Although the effects of parenting can operate through the effects on early language, they may not necessarily affect language skills later on to the same extent. However, early language skills are also a confounder of the later parenting and later language skills in the sense that parents would tend to adjust to their child's development, further enhancing what they perceive is important for their child or refraining from certain practices or investments they deemed to be inappropriate (Brandone et al. 2006).

The interventional approach cannot distinguish between those possibilities, and therefore, for a decomposition of these effects, we are left with at least three possibilities: we can either include early language skills within the mediating pathway corresponding to parenting, or exclude it and adjust for early language as yet another exposure-induced confounder of the mediator-outcome association, the third alternative been to simply ignore it. Here I present results for the former alternative and exclude early language from the parenting mediator set, while including it amid the exposure-induced confounder set. I do this because the interventional approach asks a specific question regarding what the mediated effects on some outcome would be, excluding other potentially indirect pathways. By including the language skills within the mediator set, we would be considering a hypothetical intervention that affects language skills directly, for example, asking that all these pre-school children attend some language stimulating course not involving their parents. This hypothetical intervention, though plausible, would be different than one affecting parenting styles and practices, and parental investments. However, results do not change in an important way by the inclusion or exclusion of the intermediate language skills already affected by previous parenting. Finally, the interventional mediation analyses adjust for place of residence in

Germany (i.e., East vs. West Germany); whether parents own the house they live in; mother's age at childbirth; and whether child is a first born. Finally, Table S1 in *Supplementary Materials - Chapter 4* presents descriptive statistics for all adjustment variables and confounders of the exposure-outcome and mediator-outcome associations including, though not limited to, gender, birth order, premature, low birth weight, smoking or drinking while pregnant, etc.

## Results

Table 6 shows the classification of SES into four strata, which was chosen as providing the best fit based on the lowest BIC, classifying most households into the high-SES class. The NEPS-SC1 sample, from its onset, suffers from selection bias towards children from highly educated parents (Zinn et al. 2020), which is a considerable limitation of the data for this type of study. Nevertheless, the four LCA categories do suggest a good fit with the concept of SES. For example, the high-SES class is composed of fathers and mothers with a high educational level and occupational attainment, most of them in the highest two categories of the EGP scheme. These high SES parents are also in the highest income and net-wealth categories, and none of them receives welfare. The very low-SES group, at the other extreme, is composed of parents with the lowest educational and worst occupational attainments – mostly occupational categories IIIb, V, VI, VIIa, and VIIb for fathers, and IIIa, IIIb, V, VI, VIIa, and VIIb for mothers. These very low-SES parents are mostly in the lower income and lower net-wealth brackets (e.g., including negative net-wealth or debt), and a majority of them are welfare recipients, although they are a relatively small group of the sample.

On the middle range, the categories low- and medium-SES are also distinguished from the two extremes, and are different from one another. Although the medium-SES category also has high educational and

occupational attainments, of a similar level to the high-SES class, their income and net-wealth levels are, on average, lower than the levels of the high-SES parents, but higher than those of the low-SES class. This suggests that LCA has separated the group of highly educated parents with high economic and financial resources – high SES – from the highly educated parents with lower financial resources – medium SES, something that would have been neglected had we looked at the single education or occupational dimensions.

Regarding the low-SES latent class, this one differs from both the very low- and the medium-SES groups by their relatively higher income levels, but generally lower educational levels, respectively. In terms of occupational attainment, the low-SES group is composed of a large group of technical/applied or civil servant workers, especially fathers, but a higher percentage of mothers who are out of the labor force compared to any of the other socioeconomic groups. LCA, therefore, yielded a gradual stratification of parents that combines the information of mothers and fathers' most relevant SES dimensions, and in a theoretically plausible direction. For example, a similar gradient in the language gap is also observed when considering these four latent classes, as with the reductionist approach considering only mother's educational level (see Figure S1 in **Supplementary Material - Chapter 4**).

Table 7 shows the distribution of all mediators considered in this paper. In accordance with previous studies (Attig and Weinert 2020; T. Linberg et al. 2019), parents in the high-SES group have, on average, higher levels of parental investment, more frequent parenting practices, and a slightly higher presence of sensitive mothers than parents in the very low- and low-SES groups, though the differences are really small in terms of the indicators of parenting style. Keeping in mind those differences, Figure 14 shows the size of the  $TE$ , the  $rPNDE$  and the  $rTNIE$  for the counterfactual reference

group of being from a high-SES background, at three different time points: when children were between 25-27 months of age - ELFRA-2P score -, when children were between 37-39 months of age - PPVT-4 score, and when children were 5 years of age - PPVT-4 score. Results show the well-known finding that children from very low-SES and low-SES families are the furthest away from the high-SES peers in terms of language skills at these three time points (the *TE*). These are all substantial, large effects, as found by Skopek and Passaretta (2020). The gaps are largest when children were between 25-27 months of age, smaller though persisting still when children were 37-39 months, and slightly increasing when children are five years of age. There are also similar gaps for the contrast between low-SES and high-SES groups, though these are notoriously smaller than those against the most economically disadvantaged group in this sample. There are no differences in language skills when comparing the medium and high-SES groups, which speaks to the little importance of the further economic resources of the high-SES parents, and in particular the higher occupational achievement of high-SES fathers - though, again, these effects cannot be empirically distinguished.

The main empirical findings refer to the decomposition of the total effect of SES into the direct and indirect components, also shown in Figure 14. First, the sizes of the indirect effects, measured by the *rTNIE* estimates, are always smaller than the estimates of the direct effect - the *rPNDE* - regardless of the time point, which suggest most of the SES effect operates through other pathways not involving parenting. In other words, if we were to equalize the parenting of families from different SES, we would reduce the differences in their children's language skills, but these would still remain highly stratified. This effect, however, depends on the group we are comparing the high-SES children to, and on the age of the child at which we evaluate the mediating share. For the gaps against children of very low SES

parents, the size of the indirect effect is small, and thus a small share of the SES is being mediated by parenting – When children are 25-27 months this share is  $rPM = 0.27$  with  $CI : [0.10 - 0.44]$ ; when they are 37-39 months old, the share is  $rPM = 0.18$  with  $CI : [-0.28; 0.67]$ ; and when they are 5 years or not applicable because indirect and direct effects go in opposite directions, though these are very small. For the Low-SES, the mediated share is slightly higher than for the very low SES parents – the share is  $rPM = 0.42$  with  $CI : [0.15; 0.98]$  when low SES children are 25-27 months; and, when children are 37-39 months old, the share is  $rPM = 0.21$  with  $CI : [-0.07; 0.66]$ . For the medium SES, given that the gap is rather small or non-existent, the mediated effect is null or not applicable. Thus, when children reach 37-39 months of age and even later still at five years of age, the mediated effect would be substantially smaller for all comparison groups, despite some smaller gaps being present. Therefore, these results suggest the mediated share tends to become smaller over time.

## Discussion

Overall, results suggest parenting mediates a rather small share of the effect of SES on the language skills of preschool children, especially of the later gaps when children are 5 years old. The fact that the share explained by parenting mediators reaches at most approximately one-third of the total effect of SES on children’s language skills strongly suggests that, though other pathways might be causing these inequalities, parenting could play an important though limited role in reducing inequalities to some extent. Although children from very low-SES families are the furthest away in terms of language skills from their high-SES peers, a comprehensive population parenting intervention – one that would equal the distribution of parenting done by these two groups of parents – would have a rather narrow and limited effect on reducing these gaps.

This paper contributes to the increasing breadth of evidence documenting social inequalities in language skills among pre-school children by parents' SES. Although some of the SES effect on children's early language skills is mediated, in one way or another, through parenting, the largest part of the SES effect operates through alternative pathways. Seemingly though, these alternative mechanisms do not operate through parenting and therefore would not be subject to "improvements" in the parenting skills, styles, practices, or parental investments of low-SES parents, as could be expected from previous findings too (Milkie, Nomaguchi, and Denny 2015; Sullivan, Ketende, and Joshi 2013). Therefore, language skills in children would continue to be stratified by SES, despite a hypothetical narrowing of parenting gaps achieved through parenting interventions.

I have shown that inequalities in language skills are not simply the product of deficits in parenting. Previous research has criticized the conceptualization of parenting as a simplistic transferring mechanism of socioeconomic inequalities, in particular because of the middle-class assumptions and biases that are embedded in the studies of low-SES parents, all of which are conducted by highly educated researchers - some of them parents themselves (Letiecq 2019; Keller et al. 2006). In contrast, the results of this paper suggest that low-SES families might not decide to invest less in their children, nor do they engage in less "effective" types of parenting. The differences observed in the different parenting dimensions are not being directly transmitted to their children as much as the literature believes this is happening, and as it is assumed by child development interventions on parents (Shonkoff and McCoy 2021; Mayer et al. 2019).

Parents from low resource contexts face crucial time and financial intersecting constraints that limit their ability to provide their children with those extra activities that high-SES parents take for granted (Lahire 2019). The literature of parenting is, therefore, in a constant tension between

considering poor families and their parenting as unequally positioned in the social structure, and the embrace of a deficit perspective which takes, as its standard, the parenting done by middle-class families, without acknowledging the conditions that allow for such a standard to exist in the first place (Letiecq 2019). Families, instead, should be considered as *enacting* the parenting practices and styles they received as children from their parents and caregivers, or engaging in the practices that are in accordance with their socioeconomic and cultural circumstances and experiences (Dermott and Pomati 2016).

This paper highlights various avenues for future research on the effects of parenting, the limitations of intervening on the parenting of low-SES parents, and provides three main contributions to the literature on SES, parenting, and language development of children. First, the paper presents a more holistic and encompassing measure of SES that attends to the potential for confounding and measurement error in individual, unidimensional indicators, and which appropriately incorporates both parents' characteristics. This helps to reveal the important effects of a complex construct such as SES going beyond individual indicators. Second, it is surprising that the mediated share of the joint effects of parenting is not higher than what was found in this study, especially given the number of mediators considered, and the number of potential pathways that are being simultaneously examined. Previous studies had not quantified this share and, to my knowledge, this is the first study to do so. Attig and Weinert (2020), though focused on the learning environments and parenting behavior as mediators of children's language, and including many of the indicators used in this paper, employ an empirical approach that consider these dimensions as independent of one another. Although it would be useful to know the mediator-specific pathways linking SES and children's language development through specific parenting practices or investments,

and even parenting styles, the effects of this highly specific pathways remain unidentifiable under the causal mediation analysis framework. However, given that most applied programs are focused on intervening on a “parenting package” (and not individual markers of that package), the share of SES that is jointly mediated by parenting broadly understood is of practical relevance as well. Third, these findings, though partly inconsistent with the expectations and hypothesis of the parenting literature (Ulferts 2020), highlight that universalistic ideas about parenting do not fit highly individualized developmental processes (Bear and Minke 2006) – which in itself is a reminder of the high selectivity on which most of the child development literature is based on (Nielsen et al. 2017).

Following Bronfenbrenner and Morris (2007), multiple “ecologies” are responsible for children’s developmental trajectory: the individual; the immediate and extended family context, the surrounding neighborhood, and the overall community. It is therefore not surprising that the effects of parenting are not as strong as hypothesized. There are many more alternative pathways, not involving parenting, through which the effects could be generated. These pathways – involving, for example, neighborhood effects, school environments, such as student-to-teacher ratio, the quality of teaching and access to resources, as well as extended family and social network effects, etc. – would not be addressed by interventions on parenting (e.g., J. Heckman et al. 2014).

Some limitations of this paper are, however, worth noticing. In contrast to parenting time-use studies (Milkie, Nomaguchi, and Denny 2015), it was not possible to quantify the exact time that parents spend with children because such information was not captured. Related to this, I was unable to assess the quality of the time spent in the joint activities that compose the parenting practices score, and the quality of parental investments. Parental language usage was also not directly observed. One partial explanation of

the results of this paper, therefore, is that mediators were measured with error, something that should be better addressed in future studies using time-use data and improved measurements that capture quality as well as quantity. The focus was, however, on activities that parents do with their children, as well as their reported frequency. This limits the extent to which parenting is being captured by the indicators in the study, but it may also address some aspects of parenting that are not quantified when employing time use data. Parental expectations on their child's course of development may also interact with some of these parenting dimensions. Parents' own childhood experiences of received parenting further influence how they themselves attend to their children, yet this, as well as other potentially relevant confounders (e.g., genetics and epigenetics), remain unobserved. The NEPS SC1 was not able to capture child neglect and abuse (e.g., spanking), which are particularly hard to measure, though it may have strong effects on children's language development (Widom 2014). All of these factor may further bias the estimates of the mediated share in hard to assess directions. Finally, these analyses should be replicated using a sample that has more very low and low SES households. Especially the very low SES families in the sample were under-represented. My results indicate this is a particularly vulnerable group that requires more attention from research, perhaps employing specialized studies.

A rather unexplored aspect of social inequalities in language development is inequalities by children's migration background (B. Becker 2011). Studies on US samples show that SES and racial-ethnic background of parents affect parenting and the effects of parenting on language development in early childhood (Pungello et al. 2009; Bornstein 2012; Keller et al. 2006).

However, these studies often neglect the bias present in language standardized assessments against children from minority groups (Stockman 2000; Leśniewska, Pichette, and Béland 2018). Although this topic in the

German context certainly deserves more attention (Nauck and Lotter 2015), it is challenging to study potentially culturally sensitive parenting effects on language skills inequalities by children's SES and migration background when language assessments are sensitive to the language spoken at home. Moreover, even though the effects of parenting on children are not disputed, there is still debate as to the potential for these effects to be heterogeneous, as suggested by qualitative research (Conger, Conger, and Martin 2010). Recent discussions around the effects of mother-time on children's development have called for alternative explorations of this topic in some or all of these parenting dimensions (Milkie, Nomaguchi, and Denny 2015; Nomaguchi, Milkie, and Denny 2016; Waldfogel 2016; and Kalil and Mayer 2016).

Unless the unmediated share is directly addressed by parenting interventions, the gaps will persist through other indirect pathways not affected by parenting (Manstead 2018). The specific details on how hypothetical parenting interventions could look like in practice should, in any case, depend on sound evidence-based research (Ulferts 2020). Furthermore, a specific dimension of parenting may be more important than the others. If not all that parents can do matters equally for children's language skills, what specifically could make a difference? Parenting style was one factor thought to have a substantial significance, but, besides the difficulties associated with defining and measuring this concept, it is unclear whether parenting style is something that *emerges together with* different parenting practices or parental investments, a result and not a cause. Further research should be devoted to closely examining the child rearing behavior of German parents from different SES, attending to the variability in the range of parenting behaviors, and also inquiring why parents do certain things with their children. As highlighted by Shonkoff and McCoy (2021), elucidating the complex mechanisms generating differences in early

language skills, and developing sustainable early childhood interventions, will require tackling the empirical and methodological challenges posed by parenting as a mediating mechanism.

For the case of language skills, the results of this paper would suggest, in any case, that we should consider the role of the larger context. For example, environmental interventions in combination with interventions in parenting, might be more suited to truly reduce inequalities by SES. This would require public policy to move beyond the deficit perspective on parenting, which seems unaware of the limited impact that psychology-based interventions would have on overcoming social inequalities (Cina et al. 2006). This paper highlights the potential for extending the purview on the mediation of the SES effect beyond parenting, which, although important and consequential for language development as shown here, may not suffice to reduce social inequalities. The results of this paper should hopefully encourage research on those other neglected pathways that do not involve parenting but may hold the promise of overcoming early childhood inequalities.

## **Tables and Figures**

Table 5: Description of variables used in Chapter 4

Category	Indicator	Operationalization	Waves
Parenting style	Sensitive to non-distress	The original scale of this individual items was on 1 (=not at all characteristic) to 5 (=very characteristic) scale. I created a binary indicator for each parenting marker if mother's behavior was coded as partly, rather, or very characteristic, with zero otherwise for the not at all and rather no characteristic.	1st
	Intrusiveness	"	1st
	Detached	"	1st
	Stimulating	"	1st
	Positive Regard	"	1st
	Emotionality	"	1st
Parental Invest-ments	Set of care givers who care for one or more children outside of the parents' household. This includes nursery, day care, au-pair, grandparents, other relatives or friends, or other care.	I created a binary indicator capturing whether there was any care provided by people different than the mother or father of the child. For each type of external care, a yes or no answer was available.	2nd, 3th, 4th, 5th and 6th
	Have you taken parental leave for the child since child's birth?	This question is asked for both parents of the child and at two points in time, and the response options were yes or no. These were coded as a binary indicator	2nd and 4th

Table 5: Description of variables used in Chapter 4 (*continued*)

Category	Indicator	Operationalization	Waves
Parenting activities	Child's participation in courses or groups such as toddler or play groups, baby swimming, music groups, etc.	The response options were yes or no and were coded into a binary indicator.	1st and 3rd
	Looked at picture books together; Playing together with an object which child can pull, push or purposefully grab and hold onto; Played together in or even with water; Playing together with dolls, stuffed animals, animal figurines or similar items; Playing together with building blocks or other things for inserting, stacking or building; Playing together with an item that makes noise; Interacting with child, singing, telling or showing something; Romping, cuddling or simply fooling around with child; Gone out together to enjoy the fresh air	Each of these activities was evaluated on a scale from 1 (=Several times a day) to 5 (=Not at all) and ask whether parents or anyone in the household engage with the child in any of these activities. I added them up to create a sum indicator of the intensity or frequency with which all these practices are performed with the child. The scale was then inverted so that higher values convey a higher frequency of activities and then standardized.	1st

Table 5: Description of variables used in Chapter 4 (*continued*)

Category	Indicator	Operationalization	Waves
	Reading to child or looking at picture books; Show letters to child when looking at picture books or something similar; Practicing individual numbers or counting with child; Teaching child poems, children?s rhymes or songs; Painting, drawing, or crafting with child; Reenacting something together with child; Go to a book shop together with child; Looking at picture books about nature with child; Talking with child about nature; Attend a museum or an art exhibition; Watch a movie at the movie theater; Attend an opera, ballet or classical music concert; Go to the theater; Attend a rock or pop concert	"	3rd
	Reading to child or looking at picture books; Show letters to child when looking at picture books or something similar; Practicing individual numbers or counting with child; Teaching child short poems, children?s rhymes or songs; Painting, drawing, or crafting with child; Reenacting something together with child; Go to a book shop together with child; Looking at picture books about nature with child; Talking with child about nature	"	4th

Table 5: Description of variables used in Chapter 4 (*continued*)

Category	Indicator	Operationalization	Waves
	Reading to child or looking at picture books; Show letters to child when looking at picture books or something similar; Practicing individual numbers or counting with child; Teaching child short poems, children?s rhymes or songs; Painting, drawing, or crafting with child; Reenacting something together with child; Go to a book shop together with child; Looking at picture books about nature with child; Talking with child about nature; Attend a museum or an art exhibition; Watch a movie at the movie theater; Attend an opera, ballet or classical music concert; Go to the theater; attend a rock or pop concert	"	5th

Table 5: Description of variables used in Chapter 4 (*continued*)

Category	Indicator	Operationalization	Waves
	Child engage in picture books, letter games and the like; Comparing, sorting and collecting things and the like; Playing number games, dice and the like; Doing puzzles and the like; Building and construction games; Doing handicrafts, painting, pottery and the like; Roleplaying, puppet plays, and the like; Practice sports activities, motoric games and the like; Make music, singing, dancing and the like; Experiences nature, gardening and the like; Read aloud to child at home; Tell child stories at home or look at picture books together; Show child individual letters or the ABC, for example when looking at picture books; Practicing numbers or counting with child; Teach child little poems, nursery rhymes or songs; Paints, draws or does arts and crafts with child at home; Go to a library with child	"	6th
Parents' Cultural capital activities	Parents read on a normal work day in your spare time; Number of books at home ; Books of classical literature at home; A dictionary at home; A book with poems at home; A library card at home; Works of art such as paintings at home	Each of these items were binary indicators. I added them up to create a simple additive measure of parents' cultural capital which may influence children's behavior, although not directly "parenting", it may work in indirect ways with the child observing their parents behavior and the things around them.	3rd, 4th and 5th

Table 5: Description of variables used in Chapter 4 (*continued*)

Category	Indicator	Operationalization	Waves
Confounders of parenting - language affected by SES	Premature birth	This is a binary indicator of whether the child was born before the ninth month of pregnancy.	1st
	Smoke during pregnancy	A binary indicator based on mother's self-report of whether she smoked at all during pregnancy.	1st
	Drank during pregnancy	A binary indicator based on mother's self-report of whether she drank alcohol at all during pregnancy.	1st
	Low birthweight (<2500g)	A binary indicator of whether the child was born with low birthweight, meaning less than 2500 grams.	1st
	Number of months the mother breastfed	The number of months that the mother self-reported to have breastfed her child during first year of life.	1st

Table 5: Description of variables used in Chapter 4 (*continued*)

Category	Indicator	Operationalization	Waves
	Postpartum depression	This was captured by a question on how often the mother felt depressed or sad in the last four weeks. The options given were never, seldom, sometimes, often, and always. I created a binary indicator taking the value of one when the mother responded that sometimes, often or always, with zero otherwise.	1st
	Working hours of mother and father	At each wave, the survey captured the number of working hours of both father and mother. I included an adjustment for these	1st, 3rd, 4th, 5th, and 6th
	Children's behavioral problems	This is measured by the Strength and Difficulties Questionnaire (SDQ), a standardized assessment of children's behavior. It consists of parental reports of varying degrees of problematic behavior. I standardized the score in each wave.	4th and 6th
Confounders	Place of residence in Germany: East vs. West Germany	Location of the household where survey was conducted. Only this geographical division is available.	1st
	Parents own the house they live in	I created this binary indicator of whether parents own the house they currently live in.	1st

Table 5: Description of variables used in Chapter 4 (*continued*)

Category	Indicator	Operationalization	Waves
	Mother's age at childbirth	Age of the mother when child was born computed from the difference between the birthdate of the child and birthdate of the mother.	1st
	First born	An indicator whether this was a first or higher order child.	1st

Table 6: Composition of Socioeconomic Status Latent Classes by Mothers and Fathers' Characteristics, NEPS SC1

	All	Socioeconomic Status Latent Classes			
	n (%)	Very Low (N = 221)	Low (N = 553)	Medium (N = 261)	High (N = 857)
<b>Mother's educational level</b>					
No degree or vocational/voluntary degree	116 (6)	93 (42)	22 (4)	0 (0)	1 (0)
Technical/applied or Civil Servant	448 (24)	101 (46)	263 (48)	21 (8)	63 (7)
Technical Degree (Fachhochschulreife)	528 (28)	25 (11)	247 (45)	61 (23)	195 (23)
University Education	800 (42)	2 (1)	21 (4)	179 (69)	598 (70)
<b>Father's educational level</b>					
No degree or vocational/voluntary degree	107 (6)	80 (36)	23 (4)	0 (0)	4 (0)
Technical/applied or Civil Servant	550 (29)	110 (50)	343 (62)	38 (15)	59 (7)
Technical Degree	401 (21)	23 (10)	171 (31)	39 (15)	168 (20)
University Education	834 (44)	8 (4)	16 (3)	184 (70)	626 (73)
<b>Mother's EGP occupational class</b>					
I	463 (24)	5 (2)	39 (7)	89 (34)	330 (39)
II	692 (37)	21 (10)	168 (30)	119 (46)	384 (45)
IIIa	210 (11)	78 (35)	87 (16)	12 (5)	33 (4)
IVa and IVb	37 (2)	8 (4)	13 (2)	7 (3)	9 (1)
IIIb, V, VI, VIIa and VIIb	168 (9)	70 (32)	80 (14)	4 (2)	14 (2)
Unemployed/OLF	322 (17)	39 (18)	166 (30)	30 (11)	87 (10)
<b>Father's EGP occupational class</b>					
I	700 (37)	7 (3)	43 (8)	103 (39)	547 (64)
II	505 (27)	18 (8)	123 (22)	101 (39)	263 (31)
IIIa	76 (4)	12 (5)	42 (8)	8 (3)	14 (2)
IVa and IVb	65 (3)	5 (2)	30 (5)	13 (5)	17 (2)
IIIb, V, VI, VIIa and VIIb	531 (28)	172 (78)	312 (56)	31 (12)	16 (2)
Unemployed/OLF	15 (1)	7 (3)	3 (1)	5 (2)	0 (0)
<b>Household monthly income</b>					
[100,1.100]	362 (19)	210 (95)	76 (14)	74 (28)	2 (0)
(1.100, 1.550]	435 (23)	11 (5)	276 (50)	124 (48)	24 (3)
(1.550,2.080]	552 (29)	0 (0)	178 (32)	63 (24)	311 (36)
(2.080,16.200]	543 (29)	0 (0)	23 (4)	0 (0)	520 (61)
<b>Welfare Recipient</b>					
Yes	180 (10)	164 (74)	3 (1)	12 (5)	1 (0)
No	1,712 (90)	57 (26)	550 (99)	249 (95)	856 (100)
<b>Household net-wealth Categories in Euros</b>					
(-400000,500]	444 (23)	176 (80)	139 (25)	91 (35)	38 (4)
(500,50'000]	530 (28)	32 (14)	207 (37)	139 (53)	152 (18)
(50000,170'000]	442 (23)	10 (5)	119 (22)	21 (8)	292 (34)
(170'000,150'000.000]	476 (25)	3 (1)	88 (16)	10 (4)	375 (44)

Note: NEPS-SC1. Own calculations.

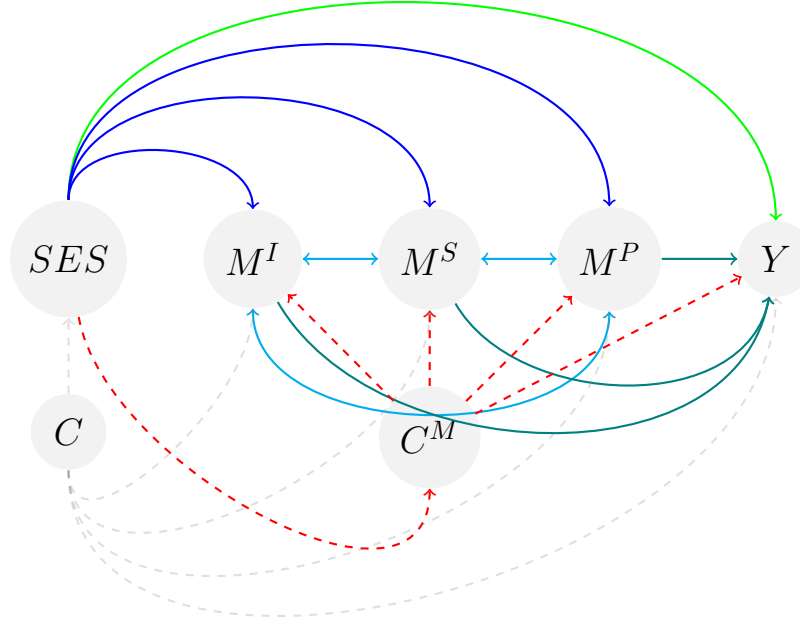
Table 7: Parenting mediators and exposure-induced confounders by SES, NEPS SC1

	All	Socioeconomic Status Latent Classes				P-value
	n (%)	Very Low (N = 221)	Low (N = 553)	Medium (N = 261)	High (N = 857)	
<b>Parenting practices score wave 1</b>						
Mean (sd)	0.00 (1.00)	-0.21 (1.05)	-0.03 (0.98)	-0.03 (0.98)	0.15 (0.97)	
Median	0	-0.3	0.0	0.0	0.3	1.7e-11
<b>Parenting practices score wave 3</b>						
Mean (sd)	0.00 (1.00)	-0.18 (1.08)	0.00 (0.99)	0.03 (1.00)	0.08 (0.95)	
Median	0.1	-0.1	0.1	0.1	0.1	9.0e-04
<b>Parenting practices score wave 4</b>						
Mean (sd)	0.00 (1.00)	-0.29 (1.12)	-0.04 (1.00)	0.16 (0.93)	0.13 (0.92)	
Median	0.1	-0.1	0.0	0.1	0.1	3.8e-13
<b>Parenting practices score wave 5</b>						
Mean (sd)	-0.00 (1.00)	-0.21 (1.05)	-0.01 (0.99)	0.11 (0.96)	0.08 (0.98)	
Median	0	-0.2	0.0	0.1	0.1	3.9e-04
<b>Parenting practices score wave 6</b>						
Mean (sd)	0.00 (1.00)	-0.43 (1.16)	-0.06 (0.98)	0.20 (0.92)	0.21 (0.86)	
Median	0.1	-0.4	0.1	0.3	0.2	7.9e-17
<b>Parents' cultural practices wave 3</b>						
Mean (sd)	8.47 (2.98)	7.19 (2.38)	7.98 (2.59)	9.24 (3.23)	9.33 (3.16)	
Median	8	7	7	9	9	1.1e-45
<b>Parents' cultural practices wave 4</b>						
Mean (sd)	11.48 (3.14)	10.40 (3.53)	10.87 (3.03)	12.51 (2.96)	12.24 (2.74)	
Median	11.3	10.0	10.5	12.0	12.0	2.7e-32
<b>Parents' cultural practices wave 5</b>						
Mean (sd)	20.62 (3.60)	18.76 (3.62)	20.37 (3.43)	21.36 (3.39)	21.59 (3.38)	
Median	21	19	21	22	22	3.2e-34
<b>Parenting style: sensitive to nondistress at wave 1</b>						
Characteristic	3,348 (98)	627 (95)	1,093 (99)	456 (98)	1,172 (99)	
Not Characteristic	70 (2)	31 (5)	16 (1)	8 (2)	15 (1)	1.9e-28
<b>Parenting style: intrusiveness at wave 1</b>						
Characteristic	142 (4)	43 (7)	54 (5)	18 (4)	27 (2)	
Not Characteristic	3,276 (96)	615 (93)	1,055 (95)	446 (96)	1,160 (98)	7.4e-14
<b>Parenting style: detachment at wave 1</b>						
Characteristic	31 (1)	11 (2)	10 (1)	2 (0)	8 (1)	
Not Characteristic	3,387 (99)	647 (98)	1,099 (99)	462 (100)	1,179 (99)	2.6e-03
<b>Parenting style: stimulating at wave 1</b>						
Characteristic	2,037 (60)	317 (48)	647 (58)	290 (62)	783 (66)	
Not Characteristic	1,381 (40)	341 (52)	462 (42)	174 (38)	404 (34)	3.0e-10
<b>Parenting style: positive Regard at wave 1</b>						
Characteristic	2,641 (77)	428 (65)	855 (77)	374 (81)	984 (83)	
Not Characteristic	777 (23)	230 (35)	254 (23)	90 (19)	203 (17)	4.4e-13
<b>Parenting style: emotionality at wave 1</b>						

Table 7: Parenting mediators and exposure-induced confounders by SES,  
NEPS SC1 (*continued*)

	All	Socioeconomic Status Latent Classes				P-value
	n (%)	Very Low (N = 221)	Low (N = 553)	Medium (N = 261)	High (N = 857)	
Characteristic	1,799 (53)	265 (40)	571 (51)	265 (57)	698 (59)	3.9e-15
Not Characteristic	1,619 (47)	393 (60)	538 (49)	199 (43)	489 (41)	
Parental investments wave 1						
Yes	1,903 (56)	124 (19)	603 (54)	284 (61)	892 (75)	9.3e-105
No	1,515 (44)	534 (81)	506 (46)	180 (39)	295 (25)	
Parental investments wave 3						
Yes	2,400 (70)	267 (41)	782 (71)	340 (73)	1,011 (85)	1.4e-55
No	1,018 (30)	391 (59)	327 (29)	124 (27)	176 (15)	
Mother took parental leave during first year						
Yes	2,619 (77)	322 (49)	892 (80)	365 (79)	1,040 (88)	2.2e-42
No	799 (23)	336 (51)	217 (20)	99 (21)	147 (12)	
Father took parental leave during first year						
Yes	1,263 (37)	67 (10)	356 (32)	231 (50)	609 (51)	2.3e-43
No	2,155 (63)	591 (90)	753 (68)	233 (50)	578 (49)	
Mother took parental leave during first two years						
Yes	781 (23)	169 (26)	273 (25)	103 (22)	236 (20)	4.2e-01
No	2,637 (77)	489 (74)	836 (75)	361 (78)	951 (80)	
Father took parental leave during first two years						
Yes	355 (10)	31 (5)	108 (10)	60 (13)	156 (13)	8.0e-04
No	3,063 (90)	627 (95)	1,001 (90)	404 (87)	1,031 (87)	
Child was cared for by others wave 2						
Yes	1,240 (36)	119 (18)	377 (34)	199 (43)	545 (46)	2.7e-11
No	2,178 (64)	539 (82)	732 (66)	265 (57)	642 (54)	
Child was cared for by others wave 3						
Yes	2,054 (60)	205 (31)	637 (57)	330 (71)	882 (74)	4.7e-29
No	1,364 (40)	453 (69)	472 (43)	134 (29)	305 (26)	
Child was cared for by others wave 4						
Yes	2,273 (67)	263 (40)	723 (65)	353 (76)	934 (79)	4.9e-11
No	1,145 (33)	395 (60)	386 (35)	111 (24)	253 (21)	
Child was cared for by others wave 5						
Yes	2,324 (68)	269 (41)	732 (66)	341 (73)	982 (83)	4.1e-04
No	1,094 (32)	389 (59)	377 (34)	123 (27)	205 (17)	

Note: NEPS-SC1. Own calculations.



Note: This diagram shows the parenting mediated share (blue lines) of the total effect of  $SES$ . The green line shows the direct effect. Interventional analysis decomposes the total effect of  $SES$  going through the different parenting mediators by properly adjusting for the confounders  $C$  and  $C^M$  which are affected by  $SES$  (i.e., exposure-induced confounding). The edges between mediators are shown as going in both directions, suggesting that relations could go both ways. However, no single pathway can be identified without making strong assumptions.

Figure 13: DAG representation of the theoretical associations between parenting mediators and exposure-induced confounding

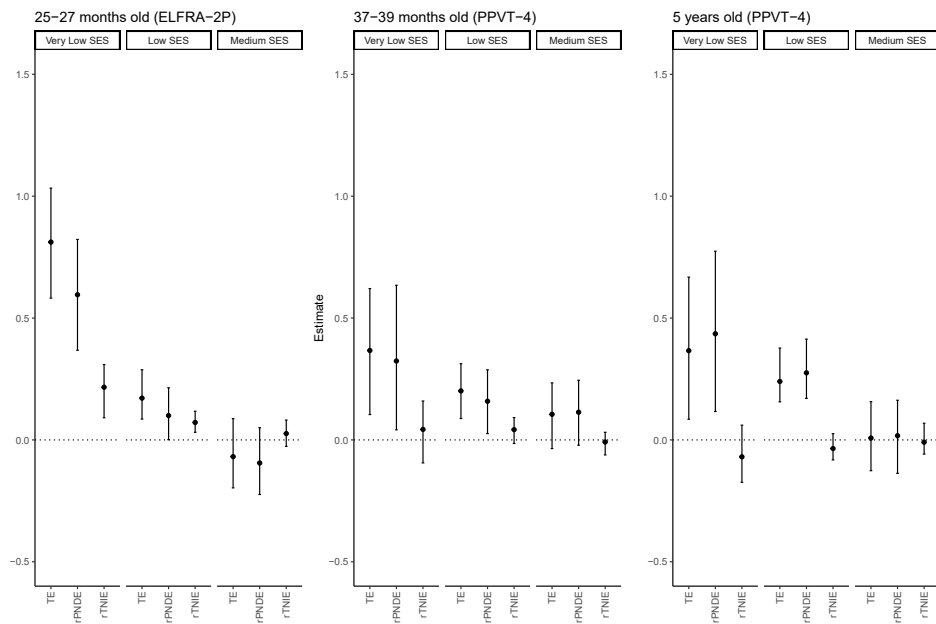


Figure 14: Interventional/Randomized mediation effects by SES on language skills at three time points with respect to High SES parents

## CHAPTER 5 - Is it just noise? Measuring unobservable cognitive abilities in early childhood

*This paper has been published as: Rodríguez Sánchez, A. (2020). Is it just noise? Measuring unobservable cognitive abilities in early childhood. Personality and Individual Differences, 166, 110162.*

*This is the accepted manuscript. The final version is available at <https://doi.org/10.1016/j.paid.2020.110162>.*

Despite a large number of studies showing that significant social inequalities in cognitive abilities can already be observed among children (Noble et al. 2015; Greg J. Duncan, ZiolGuest, and Kalil 2010; Farah 2017), little is known about how such estimates might be affected by measurement error. This is illustrated by the reliance on and the common use of the *Standards for Educational and Psychological Testing* (Association, Association, and Measurement in Education 2014), which have become the almost exclusive guideline for the choice of instruments for the measurement of unobservable constructs, but which do not say much about how measurement error might be inherent in the use of such instruments. This is the result of the standards' focus on issues of validity and reliability, and much less so with issues of measurement error from a causal inference perspective; which is a topic that remains absent from discussions on the measurement of psychological constructs, and yet it might turn out to be of crucial significance.

There are two forms of error in standardized assessments which this paper argues are of crucial importance for social science research: one refers to underlying assumptions about the types of scales which remain untested and what be sad with them, and the other one concerning measurement invariance for social groups of interest. Given that a correct quantification of unobservable cognitive abilities is required for establishing which

mechanisms drive these social inequalities, and which interventions may overcome them, this omission is a matter of concern. The operations designed to “capture” these unobservable constructs through the use of standardized assessments do not reflect the true complexity behind an adequate quantification of unobservable variables. This paper argues that measurement error affects the quantification of social inequalities in unobservable cognitive abilities, which may compromise the results of previous studies, more than what researchers might be willing to admit.

A study which took these concerns seriously was Bond and Lang (2013), in which careful consideration of the lack of cardinality of test score data made salient the need to establish whether scales for unobservable constructs are interval or ordinal, or if they at least allow for the classification of individuals into nominal groups. Moreover, these authors stressed out that there is no statistical or biological reason to assume the distribution of an unobservable trait follows the “bell-curve”. However, despite this assumption playing a critical role in the development of psychometric measurement models (Michell 2008), it remains untested. If an ordinal scale is treated as interval, however, many effects in empirical studies may disappear when they had been found before or may appear when they were not there to begin with, and they may even invert from one direction to the other as shown in Liddell and Kruschke (2018). To address this, this paper considers additive conjoint measurement checks (Domingue 2014), which are an empirical test of whether scales built from test score data conform to the properties of an interval scale.

Another concern related to measurement error is measurement invariance when comparing groups. In test development, and as part of the validation of standardized assessments, it is crucial to establish the degree to which standardized tests are sensitive to underlying differences between social groups, which are unrelated to the unobservable construct of interest.

Norming samples and pilot studies are used to accomplish this. However, validation and measurement invariance are in constant tension in the sense that, in any given application of a standardized assessment, a set of items may display differential item functioning (DIF), which, if it did occur, would crucially limit the ability of the resulting test score to establish comparisons. For this reason, this paper considers measurement invariance as a potential form of endogeneity that induces bias in the estimation of causal effects of targeted interventions or of socioeconomic characteristics ( $Z$ ) on children's unobservable cognitive abilities ( $Y^\theta$ ). Although in part discussed in the causal inference literature (Kuroki and Pearl 2014), this problem may fall into the rubric of measurement error in the dependent variable, which is usually assumed away in classic econometrics textbooks, but which might be a more serious problem when it comes to unobservable cognitive constructs.

The purpose of this study is to critically examine the extent to which standardized assessments possess the statistical properties they are often only assumed to have. The concept of measurement error is addressed from an empirical point of view, following a pragmatist and social-constructivist perspective (Baird et al. 2017). In addition to the current validation practices for psychological constructs, the paper makes use of non-parametric psychometric and representational measurement theories. To illustrate the problems of measurement, I make use of a standardized mathematics test and of two standardized psychological tests of language ability taken by a cohort of German children. The paper seeks to establish whether the use of standardized assessments as measures of underlying cognitive abilities and their use as explanatory factors in social stratification research, are justified; or whether instead, systematic error might be driving a large part of the often found social inequalities in such constructs.

## Background

### Pitfalls of the traditional validation framework

Two interrelated problems can be said to be present in standardized assessments. These characteristics may make the quantification of social inequalities in cognitive abilities more difficult than presumed by previous research. The first problem is that causal understandings of validity may not be compatible with the traditional validation framework. Contemporary notions of validity claim that “A test is valid for measuring an attribute if variation in the attribute causes variation in the test scores” (Borsboom, Mellenbergh, and Heerden 2004, 1067); which for cognitive tests would imply that changes in brain structure and function at a neuronal level, for example, correspond to higher cognitive ability. Therefore, changes “in the brain” (e.g. as a result of development or targeted interventions) should cause nonnegative changes in test scores, if researchers wish to determine that a standardized test is valid, i.e., that it measures something happening to the brain (e.g., more neural connections leading to higher mathematics ability, similar to how in a blood cell count test the presence of disease causes changes in the cell count).

In the psychometric framework of item response theory (IRT), cognitive abilities are presumed to lie on a  $[min, max]$  range. Mathematics ability for example would fall into such types of hypothetical ranges, which should somehow capture how the structures and functions of the brain indicative or supportive of a more mathematically-skilled child are distributed in a population. Test scores of mathematics skills, which have their arbitrary scales according to the number of items and to how items are graded, should map onto the hypothetical range of cognitive abilities in a nondecreasing functional relation (Vautier et al. 2012). Such evidence, however, has not been provided by the test developers of any of the current existing cognitive

tests. Besides from evidence based on brain imaging suggesting infants possess fundamental cognitive capacities to differentiate auditory and visual stimuli (Dehaene-Lambertz and Spelke 2015), standardized tests aimed at “measuring” cognitive abilities use the verb measure in a metaphorical sense (Briggs 2013), and are completely disconnected from the concepts of measurement in the natural sciences (Uher 2020). This suggests that interpretation of test scores might be hampered by a fundamental lack of validity when the measurement operation is considered as a unified scientific framework.

And even though no unique concept of validity exists (Newton and Baird 2016), validity and measurement, in general, are still downplayed in the social sciences, where test scores are used to show the existence of gaps between social groups and to explain such gaps. As stated before, present validation state-of-the-art methods follow the *Standards for Educational and Psychological Testing* (Association, Association, and Measurement in Education 2014), in which a combination of exploratory or confirmatory factor analysis (EFA or CFA); internal reliability estimates such as Cronbach’s  $\alpha$ ; and correlation coefficients between test scores and other relevant outcomes, are used to assess the internal consistency of latent variables. Validation of scales of educational constructs follows in turn a similar framework (Shear and Zumbo 2014, 91–111).

However, this understanding of validity might be misleading, and reliability therefore much harder to achieve. As shown by Maul (2017), and earlier by Wood (1978), CFA or IRT models do not provide researchers with the means of detecting a truly underlying structure in data from standardized assessments. These numerical procedures are not designed to falsify the hypothesis that an underlying unobservable construct is driving the correlations, because almost always these methods will find some underlying “structure” when applied to data. They are, after all,

dimensionality-reduction techniques based on correlation, not causation. Moreover, more specialized procedures, such as Rasch models and its developments, have been shown to provide appropriate fit even when their assumptions in simulation studies are violated (Karabatsos 2001, 394–95; and for estimating “coin-flipping” ability Wood 1978). Even though no perfect fit is expected from such models, these multiple laboratory and simulation studies have brought forward the disconnection between validation as causal and validation as model fit. The construct of “gibberish” can be measured with high reliability following the standard approaches because the numerical procedures within the current standards are not connected to the most intuitive definition of validity, namely that a standardized test measures what is supposed to measure and nothing else (Borsboom, Mellenbergh, and Heerden 2004). Therefore, one type of error in which social sciences research might have been incurring on is to take test scores as indicators of an underlying construct, when in fact test scores might conflate the operational definition of a given construct (i.e., items on a test related to cognitive ability) with the construct of the cognitive ability itself. [Uher (2020); p.995]. In most cases of cognitive constructs, in fact, items define the construct, and not the other way around (Baird et al. 2017, 324)

### **Ordinal v. Interval: properties of scales**

Second, there is the problem of the scale at which constructs can be supposedly “measured”. Three types of scales are often in mind: nominal, ordinal, interval, or ratio (Velleman and Wilkinson 1993). This issue is still a matter of dispute, but the effects of considering underlying constructs as ordinal or metric may introduce various kinds of errors in empirical studies. O’Brien (1985) provided an overview of the problems which result from treating ordinal variables as if they were interval ones. Doing so may lead to

several misleading results in the estimation of effects, such as the prediction of values below or above the scale's range, the lack of variability in response options, ceiling, and floor effects, among other well-known errors (Agresti 2012, 5–7). However, more recently Liddell and Kruschke (2018) showed that more severe errors may occur when incurring in this practice: false alarms, failure to detect effects, and even inversion of effects.

Relatedly, though different from this, is the discussion in Bond and Lang (2013) about the salience of the assumptions underlying the scoring process of mathematics achievement tests in the U.S., which had been discussed in the value-added assessment literature as well (Ballou 2009). If test scores only possess ordinal properties, then different assumptions about the unknown distribution of the unobservable latent construct can lead to contradictory results. In fact, dramatic changes arise in the direction of effects by assuming different distributions for the unobservable construct of mathematics ability, which can be done by applying nondecreasing monotonic transformations to the test-scores (Bond and Lang 2013). For this reason, given that there is no evidence that unobservable constructs are quantitative attributes; nor that these unobservable constructs have been measured on an interval scale, research should empirically establish which qualities are 'measurable', i.e. behave as quantitative magnitudes, orders, or classes of a hypothesized construct.

### **Measurement invariance as measurement error**

The third related problem is more general, but perhaps more severe than issues of validity and scaling. Serious epistemological criticisms of psychometric measurement models abound, and an overview of literature discussing psychometrics as an entirely or partly pathological science, as claimed by Michell (2008), can be found in the works of Johnson (1936), Michell (2008), Trendler (2013), Humphry (2013), Mari et al. (2017), Maul,

Iribarra, and Wilson (2016), Briggs (2013), Vautier et al. (2012) and Lacot, Afzali, and Vautier (2016), among others. Despite the little attention they have received in the applied literature, these authors point to a common feature of standardized assessments: measurement invariance. In a nutshell, their argument goes back to two characteristics of standardized tests. First, that measurement through questionnaires or tests makes explicit use of the human mind in the measurement process, i.e., a child answers questions on a standardized test. Second, that regardless of how much control is placed on the testing situation, the human mind is not a reliable measuring device (Uher 2020). For example, in standardized educational tests, correct responses allow observers to infer that a child masters a particular skill or competence. The opposite inference, namely that the child does not master the skill, cannot be drawn from an incorrect response because a wrong answer might have resulted from language barriers; unfamiliarity with the testing situation; lack of concentration; lack of motivation; lack of working memory; lack of confidence; stereotype threat; tiredness; stress, anxiety, or fear; or from multiple-way and nonlinear interactions of many of these factors (Banerjee 2016). Which are, under the assumption of unidimensionality on which most if not all psychometrically validated tests are built (Baird et al. 2017, 323–24), unrelated to the cognitive construct itself.

Beyond traditional factors accounted for in educational research (e.g., opportunities to learn as determined by the teacher, classroom or socioeconomic characteristics, cheating, guessing, etc.), these other extraneous factors which may hinder students' performance are far from randomly distributed among a population. In fact, they are likely concentrated among disadvantaged groups and cannot be factored out from the measurement operation. Thus, it is easy to derive alternative hypotheses to explain differences between social groups or changes in scores as not

unequivocally caused by corresponding differences or changes in the hypothetical unobservable attribute residing in children’s brains. This in and of itself constitutes a serious threat to the validity of these instruments for establishing differences or inferring changes in psychological or educational constructs.

Concerning this, measurement invariance is thus one of the main sources of measurement error related to the above problems. When some test displays measurement invariance, the operationalization of the construct may not be equivalent among different groups in a population. Therefore, some aspects of the test, or specific items in it, affect the response behavior in a manner not relevant to the construct. Test developers often state that pilot studies on similar samples allow for the detection and exclusion of items displaying DIF (Penfield and Camilli 2006). However, even within the traditional framework, validation is understood as an ongoing process (E. K. H. Chan 2014, 4). Therefore, a scale’s “good” psychometric properties in similar samples are never sufficient criteria to determine the validity of an instrument for comparison between groups. A test is valid if, and only if, it has been validated for well-defined purposes in the contexts of its application. Thus, no DIF should be observed each time the test is used when comparing groups.

And here is where the problem of measurement invariance may have lead researchers astray when they are interested in establishing causal associations using cognitive constructs. Basic causal effect notation may help to highlight the problem of measurement error in the quantification of unobservable attributes for causal inference, as shown in the DAG of Figure 15. A basic assumption of all psychometric models to date is that  $Y$ , the score on some standardized assessment, is an unknown function  $f$  of  $Y^\theta$ , representing the true cognitive latent construct, and of  $U^Y$  representing an error term, i.e.,  $Y = f(Y^\theta, U^Y)$ . Social scientists focus on unbiasedly

estimating effects of some characteristic or intervention program  $Z$  on the cognitive abilities or skills of children, the arrow  $Z \rightarrow Y^\theta$ . What is done in practice is to estimate the response surface  $\mathbb{E}(Y|Z)$  employing some type of model in which an additive error term, representing all other factors not included in the model, is usually added, i.e.,  $Y = Z\beta + \varepsilon$ . However, even in this simple framework, the typical assumption that the error term  $U^Y$  on the dependent variable is random or exogenous may not be warranted for the case of cognitive constructs and their operationalization (i.e.,  $\sigma_{Z,U^Y}^2 \neq \sigma_{\varepsilon,U^Y}^2 \neq 0$ ). It is hard to conceive of a standardized assessment unaffected by some of those extraneous factors captured by  $\varepsilon$  which occur in systematic ways. Therefore, because  $Z \rightarrow U^Y$  or  $\varepsilon \rightarrow U^Y$  might be causal effects affecting the measurement error component, especially if there is evidence of DIF, estimates of the effects of  $Z$  on  $Y$  might be overestimated due to what causal inference refers to as confounding. In general, if there is measurement invariance for the variable of interest  $Z$ , estimates of the size of the green arrow in Figure 15 will be biased because of measurement error in the dependent variable  $Y$  that is correlated with the independent variable  $Z$  and or the error term  $\varepsilon$ . These potential correlations, which are assumed to be zero in econometrics textbooks, should not be assumed away in the case of unobservable cognitive constructs based on standardized assessments as dependent variables. Especially so when little is known about how  $Z$ , or any of the omitted variables captured by  $\varepsilon$ , may causally affect  $U^Y$ .

Psychology’s “atomic bomb”, to coin Borsboom and Wijsen (2017, 444)’s term when referring to psychological tests’ role in the scientific world, surreptitiously fell over social science research practices without causing too much noise, but perhaps a lot of damage. Following the work of Baird et al. (2017), this paper proposes the use of readily available empirical tests that may make visible the pitfalls of present validation practices in educational research, which hide these issues. In what follows, this paper discusses these

methods, namely nonparametric item response theory and DIF; as well as on the representational measurement theory (i.e., additive conjoint measurement). By doing this researchers might be able to ground the validation process in empirically supported claims and not on untested assumptions. When the usual validation framework is connected to these additional analyses, it might be possible to overcome some of these pitfalls: First, by recognizing that these might be the cause of bias; and second, by providing appropriate warrants for the kinds of causal claims researchers would like to make about social inequalities in cognitive abilities when interpreting test scores.

## Methods

### Data

To assess the above-mentioned problems in an empirical study, I evaluated two dimensions of children’s cognitive development: mathematics and language abilities. As an example, and only to illustrate this, I chose data from the National Education Panel Study Starting Cohort Number 1 (NEPS-SC1, Blossfeld, Roßbach, and Maurice 2011)<sup>1</sup>; but I argue that the problems of measurement are inherent to all standardized assessments. The NEPS-SC1 contains a diverse set of standardized tests to measure children’s development in various dimensions. This section describes the tests’ characteristics, and descriptive statistics can be found in **Supplementary Materials - Chapter 5** Tables S7 and S8. The sample was generated using a complex random study design and consists of a cohort of newborns who were officially registered between February and July of 2012. In each

---

<sup>1</sup>NEPS-SC1 (doi:10.5157/NEPS:SC1:6.0.0). From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LifBi) at the University of Bamberg in cooperation with a nationwide network. The NEPS-SC1 sample consists of a cohort of newborns in Germany that has been followed for over seven consecutive years.

follow-up survey, children's characteristics and performance in standardized tests were recorded along with parents' characteristics.

**Measurement Instruments** First, a standardized **Mathematics test** in Wave 5 was carried when children were around five years old. This test consisted of 20 items on five different categories of mathematics competency: a) sets, numbers, and operations; b) units and measuring; c) space and shape; d) change and relationships; and e) data and chance (for information about the test and its scaling procedures see Petersen and Gerken 2018). Second, the Peabody Picture Vocabulary Test Fourth Edition (**PPVT-4 Verbal ability**) was applied at Wave 4. The PPVT-4 assesses receptive vocabulary or verbal skills in children and adolescents in the German norm-referenced sample (see for an overview of the test Roßbach, Tietze, and Weinert 2005). The test contains 228 items divided into 19 sets, each containing 12 items. These items are of varying difficulty, but the total number of correct responses is computed to establish the progression of children's language development. The PPVT-4 is also used as an indicator of cognitive ability, given that comprehension of language and concepts plays a major role in it. And third, the *Elternfragebogen für die Früherkennung von Risikokindern 2* (ELFRA-2) <sup>2</sup>, which is a standardized questionnaire containing a list of words and utterances that a child should be able to say at a given age. It is filled in by the child's main caregiver and consists of three subscales that assess productive vocabulary, syntax, and grammatic abilities in German. This paper makes use of the subscale on productive language (**ELFRA-2P**). The ELFRA-2 is used as a screening test for diagnosing delays in language development in German children aged 24 months (see for an overview of the test Grimm and Doil 2006). The scale has been shown to have adequate reliability values according to the standards but is sensitive to the language spoken at home (Sachse and Von

---

<sup>2</sup>The Parent's Questionnaire for the Early Diagnostic of Children at Risk 2

Suchodoletz 2007).

## Methods

As explained above, the two main assumptions refer to the scale of the unobservable construct and the issues of measurement invariance. There are some methods available to empirically explore the extent to which assumptions about unobservable constructs are warranted. To address the first of these, two frameworks are considered: the representational theory of measurement and the non-parametric item response theory. The first one focuses on whether the quantitative structure assumption holds in the data. From previous research, it is known that whenever the difficulty of items in a test differs, as is the case of most standardized tests, raw or standardized sum scores do not yield an interval scale (Wright 1992; Ballou 2009). Given that sum scores only provide ordinal information, which allows to rank test-takers, researchers scale their data through IRT models to obtain interval scales. Although these models are supposed to guarantee estimates of ability are continuous, it is argued whether many of the most crucial IRT assumptions are warranted, especially when no scale built from standardized assessments so far is of an interval kind (Kyngdon 2010; Domingue 2014); and when it might be that psychological attributes cannot be measured at all (Trendler 2018).

However, the hypothesis of quantifiability of an attribute can be verified by checking if a specific functional relation among the set of respondents and the set of items holds in the data (Luce and Tukey 1964). Karabatsos (2018) and Domingue (2014) developed methods to assess a stochastic version of the axioms of transitivity, antisymmetry, and strong convexity for an ordinal relation; and associativity, commutativity, monotonicity, solvability, positivity, and the Archimedean condition for an interval relation (Heene 2013); which are the core assumptions from which interval scales can be

derived from according to the ACM perspective. The connection among these conditions and the single and double cancellation axioms is elaborated in Domingue (2014). It is these conditions which to some extent can be empirically tested, although Karabatsos (2018, 324) pointed out the limitations of this approach for higher-order cancellation axioms. However, when these conditions hold, the claim that the responses to a standardized test yield an interval scale is more plausible. Violations of the quantitative structure assumption are expressed in the percentage of comparisons of adjacent  $3 \times 3$  matrices that do not comply with the single and double cancellation axioms. Although such checks have been used in some empirical applications, the use of and familiarity of researchers with ACM is far from extended within psychometrics (Domingue 2014; Dimitrov 2016). In this paper these checks are performed on the three standardized tests and compare the results obtained.

Another alternative route to examine the interval scale assumption is to consider ordinal psychometric models and their properties. By empirically checking whether these are present, one can indirectly gather evidence against scales as intervals if the ordinal relations happen not to be found in the data. These ordinal relations can in turn be tested employing Mokken scale analysis [MKS; Sijtsma and Meijer (2006)], which considers non-parametric item response models, such as the monotone homogeneity model (MHM). These might be used as an exploratory tool to study response patterns and establish whether ordinal scales can be built from data (Sijtsma and Ark 2016), even when ACM checks show violations of assumptions for the hypothesized quantitative structure. This framework allows for a test of three of the most basic psychometric properties underlying all psychometric models: unidimensionality, monotonicity, and local independence. These properties should hold in the data before considering fitting a model for ordinal scales (Sijtsma and Ark 2016), and

especially before considering fitting a model that assumes interval scales. An advantage of Mokken scales, however, is that a scale might be constructed out of subsets of items that do conform to the MHM's assumptions. Mokken subscales are those that conform to these three properties, which also underly all other IRT models. The construction of a Mokken scale is shown for the mathematics test for illustration purposes. However, only scalable items at  $c = 0.3$  are chosen, and items causing violations of monotonicity or local independence are excluded.

As for the measurement invariance assumption, the analysis was based on the framework proposed by Rosenbaum (1984, 428). The sum score  $R_{(-j),[i]} = \sum_{l=1}^J y_l$  for  $l \neq j$  for each child  $i$  excluding item  $j$ , also known as rest score, might be taken as a reduced form of the test, and might be used to empirically assess if other extraneous factors associated with some characteristics of interest are affecting the probability of correctly answering the excluded item. I estimate the relative risks for the event of correctly answering item  $j$  in the test, conditional on the score the child would have received had this item been excluded from the test, namely for the event of  $\mathbb{P}(y_j = 1 | R_{(-j),[i]}, Z)$ . This approach has a number of attractive features. Assuming this rest score is the best predictor for correctly answering the excluded item  $j$ , children's characteristics should be unrelated to the probability of a correct answer. I estimate relative risks using a log-linked binomial model for a set of sociodemographic and socioeconomic covariates: preterm, gender, migration background, socioeconomic status (SES) and language ability (for the estimation of SES see Table S10 in

**Supplementary Materials - Chapter 5**). For this step, the two polytomous items in the mathematics step were recoded as follows. If at least one out of four options was correct, the item was scored as correct, ignoring the gradation difficulty implied by the question. All other items in the test were dichotomous. Twenty models, one for each item in the test,

were estimated, and the relative risks with 95% confidence intervals are presented. After adjusting for ability measured by the rest score, relative risks of probabilities should equal one or be close to one, without showing patterns besides expected random fluctuation. If the mathematics test is fair, valid, and reliable, then the rest score should explain all variation affecting the probability of correctly answering any other item in the test.

Finally, a more conventional approach to measurement invariance is explained in Penfield and Camilli (2006), which discusses the relevance of nonparametric methods to achieve this without making unrealistic assumptions about the scale of constructs. One of these methods is based on the generalized Maentzel-Hazel (gMH) statistic, on which tests might be performed without requiring the estimation of an IRT model with its corresponding assumptions to hold (results of this statistic are found in **Supplementary Materials - Chapter 5** Table S18). If a scale does not equally evaluate individuals from different groups, then a test might be considered to be biased, and its items should be examined and potentially excluded if, on closer examination, they are found not to be measurement invariant. This is how the making of standardized assessments takes place: by considering a large set of items of which some are eliminated because they do not conform with the desired structures of a construct (Baird et al. 2017, 331). The gMH statistic was used with continuity correction and P-value adjustment for multiple comparisons by the Holm method for a comparison of more than two groups. Measurement invariance is checked for the social groups of preterm children, girls, children with migration background, and children from low-SES families. The raw score was used to match children from each group and a threshold or cut-off value score was selected to classify the corresponding items as displaying DIF (Magis et al. 2010).

However, such a meticulous analysis of the data seems insufficient to show that measurement error causes bias if one cannot show that the connection

between the assumptions and the estimation of effects. The last analysis presented in the paper, therefore, focuses on showing that the above-mentioned considerations do make a difference. To determine this, this paper compares estimates of effects of several covariates on mathematics ability based on IRT ability estimates; on the sum of correct responses; and on the sum of correct responses in a Mokken scale. The comparison considers, in addition to basic linear regression models, the cumulative link model (CLM) for the probabilities of being at different quartiles of the distribution of ability (Winkelmann and Boes 2006, 175), which would be appropriate if indeed scales are ordinal and not interval scales. Differences between the two modeling strategies are assessed descriptively. Given that variance in ordinal models cannot be decomposed as in linear models, and that estimated parameters associated with each group of covariates will change with the inclusion or exclusion of additional covariates, the model assessment was done by estimates of changes in probabilities, i.e., average marginal effects (AME). Only complete cases were used for these analyses. Finally, in each of these three analyses sample sizes differed depending on the test being assessed, but this is beyond the point and results hold when the sample of complete cases is used (results available upon request). Multiple imputation analysis was not applied because these methods make use of associations already present in the data, which the analysis presented here aims to empirically assess. Except for relative risk estimates, no inferential results are presented, and regression models are shown only for illustrative purposes, and only the univariate estimates of relative risks are estimated taking into consideration sample design. Tables S7-S8 in **Supplementary Materials - Chapter 5** show the sample sizes used for each analysis.

## Results

Results of the standard validation practices can be found in Tables S11-S16 in **Supplementary Materials - Chapter 5**. It suffices to say that these results conform to the expectations of validity and reliability according to the standards. Based on those results, and as a running example, one can consider the probability of scoring below the 25th percentile of the distribution of mathematics ability as measured by the full battery of items on the standardized mathematics test by various sociodemographic groups, which is scaled by a parametric IRT model. This group corresponds to the less able in terms of unobservable mathematics ability according to the standards. For children born preterm, the unadjusted relative risk of being among this lowest scoring group is 1.724 times that of full-term babies (C.I: [1.323, 2.208]); for girls, the risk is 0.979 times that of boys (C.I: [0.855, 1.121]); for children with migration background it is 1.736 times that of nonmigrant background children (C.I: [1.45, 2.063]); and for children of parents in the least well off socioeconomic status (SES) 3.162 times the risk than children of the most wealthy, highly educated and better employed parents (C.I: [2.569, 3.902]; own calculations).

Regarding the results which concern the scales of these constructs, results shown in Table 8 suggest several violations of the single and double cancellation conditions which should test for an ACM structure in the mathematics test responses. Both the weighted and unweighted proportion of violations are high when compared to the 2% for the unweighted and 1% for weighted violations for data simulated from a Rasch model and subjected to the same checks (Domingue 2014). Heene (2013) suggested such results would be obtained for many data sets that fit some parametric IRT model, even when violations of its assumptions are present, so this result is not surprising. However, what is surprising is that none of these standardized

tests satisfies the conditions of an interval scale (i.e., that the differences between the units on test scores are of an equal interval). These tests do not fulfill the assumptions of a quantitative attribute despite evidence in favor of appropriate fit following the *Standards for Educational and Psychological Testing*. These results are further confirmed by the Mokken scale analysis. Unidimensionality, monotonicity and local independence assumptions as well as invariant item ordering, which are shown in Table 9 and in

**Supplementary Material - Chapter 5** Table S17, suggests poor fit of this data with these ordinal model, which shows some of the items in the mathematics test are even unscalable. There was, however, evidence for a weak scale ( $\psi = 0.3$ ) in accordance to this model's properties, which can be conformed from 11 of the items in the original test. These items do conform an ordinal scale without either violation of monotonicity or local independence. However, the Mokken scale of mathematics ability does not conform to the properties of a quantitative attribute either, given that the weighted average number of violations is 11.393 and the unweighted 17.07.

Moving to the second concern regarding measurement invariance, Figure 16 presents the estimated relative risks associated with a one-unit increase in the rest score. These risks are ordered from easiest to hardest items according to the percentage of correct responses to the question in the test. One extra point in the rest score increases the probability of correctly answering the excluded item; these are, as expected, all greater than one. Moreover, the rest score's predictive ability increases with the difficulty of the items. In this sense, this result validates the use of the test in the prediction of a correct response. However, and as shown in Panels A, B, C, and D of Figure 17, other covariates also remain predictive of a correct response in the items, even after controlling for mathematics ability as measured by the rest score. For the language ability tests, ELFRA-2P and PPVT-4 shown in panels E and F, no association with the probability of

correctly answering an item is seen; but this is not the case for low-SES children. For 13 out of 20 items, the relative risks are below 1, meaning that there was a lower chance of correctly answering 13 items in the test despite controlling for mathematics ability as measured by the rest score. The more difficult items also show effects for being born preterm and also for being a girl (some in a positive direction). However, except for SES, the effects of the other covariates do not follow any significant pattern; but it is possible that this standardized test has captured something other than mathematics ability and therefore that systematic error may have permeated the measurement operation.

To confirm this, DIF in the different standardized tests shown in Table 10 presents the share of items that were flagged as presenting DIF. Results suggest DIF is present in all three of these standardized tests. Of special concern are those items for the mathematics test that show much higher odds of being correctly answered for groups different than the focal groups, as displayed in the gMh coefficients in Table S17 in **Supplementary Materials - Chapter 5**. The mathematics test has the largest share of items flagged as DIF for the different groups. Such high percentages might be indicators that the test is biased, especially concerning children with a migration background, as well as children from low-SES families. The group of preterm children does not show noticeable differences, although some of the items are regarded as presenting DIF.

Finally, in Figure 18 the standardized coefficients and average marginal effects for the different analytic strategies described before are presented. Interestingly, there are multiple differences between these analytical strategies. Using the Mokken scale as a linear scale showed larger effects for preterm and migration background, whereas the parametric IRT (PIRT) and sum score showed the largest effects were for the language ability covariates. Moreover, when considering effect sizes, as shown in Table 11,

differences between the linear regression and the CLM are noteworthy. AMEs, which are taken as overall effect sizes of the predictor variables, showed that low-SES had the largest effect of them all. By contrast, according to the linear model, it is language abilities which have the largest main effect. Furthermore, the effects of being preterm or having migrant background were larger when using the Mokken scale than when using the scale built with the full battery of items; whereas for socioeconomic status these effects were smaller when using the Mokken scale than the full scale. The most striking effect is that the direction of the effect of migration background changes comparing the PIRT to the Mokken scale. The fact that these noticeable differences were found in an easy comparative exercise highlights how drastic a study's conclusions might change when following different analytical strategies, as argued in Liddell and Kruschke (2018).

These results, partial as they are, especially for the EFLRA-2P and the PPVT-4, which were mostly because of space limitation, show that these three standardized tests do not show the properties of an interval scale, nor do they seem unbiased for the groups here chosen. The items that fit the assumptions of IRT in the Mokken mathematics ability scale, excluding items flagged as DIF for each of the four covariates here examined (i.e., items z17s, v061, z121, and r14s), would leave the mathematics scale with only 7 items on which ordinal comparisons can be safely made.

Based on those seven items, and an ordinal scale built on them, the relative risk of being among the lowest-scoring group<sup>3</sup> in the Mokken scale of mathematics ability is, for preterm children 1.476 times the risk than full-term babies (C.I: [1.198, 1.798]); for girls 0.983 times the risk than boys (C.I: [0.886, 1.09]); for children with migration background 1.307 times the risk than non-migrant background children (C.I: [1.124, 1.512]); and for

---

<sup>3</sup>Which means having correctly answered 0, 1 or 2 out of 7 mathematics questions, corresponding to the 25th percentile of children's scores on this reduced form of the test.

children of parents in the least well off socioeconomic status (SES) 2.202 times the risk than children of the wealthiest, higher educated and better-employed parents (C.I: [1.872, 2.59]). These are substantially smaller, though still important social inequalities when compared to those indicated by the full battery of items. However, these estimates are qualitatively different because they are based on ordinal information, not on a metric “measure” of mathematical ability, as presumed by the estimation of an IRT model. Such an ordinal conception of inequalities lends itself to a different and more fruitful discussion of social inequalities in unobservable cognitive constructs. Even though this alternative analysis might be favored by some and not others, it has one advantage over the standard assessment: it is based on assumptions that are supported by the data, whereas the standard analysis is not. It is interesting to note that if assumptions on which IRT scale tests are not warranted, then analysis based on those assumptions, such as logit scale based DIF, are also not warranted. These results are preferable in the sense of conforming to the assumptions of the models used to scale them, but do not represent the only alternative. Evidence thus suggest that although standardized assessments are not just measuring “gibberish” or noise, to clearly separate the signal from the noise in this type of data, and to unbiasedly estimate causal effects, much more than what the *Standards for Educational and Psychological Testing* assume is needed would be required.

## Discussion

The present study was designed to explore the extent to which two forms of measurement error might be present in data generated by standardized assessments. As hypothesized, results suggest that treating these scales as if they had interval properties may lead to erroneous conclusions that are not warranted when scales are found to only possess ordinal properties. But

also, results indicate that measurement error creates a strong bias in the estimation of the effects of typical socioeconomic and demographic covariates. Therefore, measurement error, at least in the standardized assessments analyzed in this study, is pervasive in the measurement of cognitive abilities. Future research should establish the extent to which these problems can be found in the quantification of other unobservable attributes of cognitive or “non-cognitive” domains, such as personality assessment, but other research suggests this might be the case as well (Uher and Visalberghi 2016).

What is surprising in this study is that none of the scales here considered displayed the less demanding properties of nonparametric measurement models and consequently nor those of an interval scale, even though such types or standardized tests are common in the literature and have been validated according to the standard validation framework. Previous studies observed that the scale at which unobservable constructs might be measured is far from fulfilling the strict criteria of interval scales (Domingue 2014; Karabatsos 2018). More worrisome, however, is that current validation practices hide this particular problem because the assumption of equal-interval unobservable constructs which can be then measured is merely stated, but never tested (Michell 2008). The data and measurement theories outside of psychometrics, however, suggest these attributes may not be quantifiable (Trendler 2013; Uher 2020). As mentioned in Borsboom (2006), IRT aims at modeling the interaction between a person’s ability as a latent and unobservable trait and a given item stimulus, but this theory does not guarantee that their underlying assumptions hold in any standardized assessment based on the IRT paradigm, much less in classical test theory. Therefore, a change of paradigm in the validation of standardized assessments to measure unobservable cognitive skills might be needed, one that is empirically able to test its underlying assumptions.

These results thus support the idea that the use of test scores for comparisons on a difference-scale must be warranted by the measurement operation and not by untested assumptions about what the attribute is (Velleman and Wilkinson 1993).

On the issue of measurement invariance, results indicate that low-SES and migration background seem to affect the chances of correctly answering a subset of items in a test, even after adjusting for “mathematics ability” as measured by the rest score. These results were further confirmed by the analysis of DIF, which furthermore also pointed to similar problems in the quantification of verbal ability. Two implications of this finding warrant discussion. First, teachers’ use of scores on standardized tests in classroom activities may reify social inequalities when differences of an ordinal kind are misunderstood as being of an interval one if indeed they are based on similar standardized assessments in the classroom (Dalziel 1998). In a scenario in which child A has answered 12 out of 20 questions correctly, whereas child B got 6 out of 20, and child C only 3 out of 20, the difference in mathematics ability between children B and C ( $|b - c|$ ) compared to the difference between children A and B ( $|a - b|$ ) is not twice as much ( $|a - b| \neq 2 \times |b - c|$ , where  $a$ ,  $b$  and  $c$  are taken as units of an unobservable construct, ability or competence). How far behind child C is from child B or A, or child B is from child A, remains unknown and might even be a meaningless question to ask in the first place. From a pragmatist perspective, if hypothesized changes in the latent variable cannot be empirically traced back to quantitative changes in the score produced by the standardized test, then the information provided by test scores may make comparison across groups invalid.

Second, more caution should be had when interpreting the estimates of the effects of covariates on unobservable cognitive abilities. As shown in this paper, measurement error may notoriously affect the size of inequalities

when measurement error is not considered. It is possible to hypothesize that this results from statistical confounding or selection bias, as argued above (Kuroki and Pearl 2014). Again, the reason being that it is no longer possible to discern whether differences are caused by changes in the attribute, or by changes in systematic error component, or maybe by both (Vautier et al. 2012), which would be an insidious form of endogeneity present in almost all tests that display DIF for the effect of interest  $Z$ . The results of this paper thus suggest that caution is warranted in using test scores to compare children across social groups based on difference statistics, which might be, as shown here, an unjustified use because scales might be ordinal, not quantitative, and because of measurement invariance.

The present results are important in two major respects on which future work should focus. First, social scientists have long documented the lower academic chances of children from disadvantaged backgrounds (Bourdieu and Passeron 1964). And although social stratification research suggests that data on competencies mediates the generation of such inequalities (i.e. primary versus secondary effects), it is not clear whether inequalities in education result from deficits in cognitive skills; or from standardized tests reliably measuring both signal and noise. An implication of this being that standardized tests and evaluations of children might simultaneously create and certify these social inequalities (Millet and Croizet 2016; Grodsky, Warren, and Felts 2008). Second, although this paper deals with one specific type of assessment, social science research should dedicate more effort in thinking about the distance between behavior (i.e. taking cognitive abilities as a type of behavior) and assessments that supposedly measure that behavior. The goal of building scales or subscales with expected properties is motivated on “technical” grounds, but whether a given pattern of responses is being caused by the construct being assessed, or whether the pattern results from the hypothesized and desired characteristics of a

measurement instrument, as arguably happens when selecting items in PISA (Baird et al. 2017, 331), is still unknown. These results, however, corroborate the findings of Uher and Visalberghi (2016), where analogous problems in the assessment of personality are delineated. Because a test's reliability and validity using the traditional validation framework may result from tests reliably capturing noise; and from educational institutions making extensive use of standardized tests in the classroom, it might be that "measurement error begets measurement error".

Although early childhood inequalities in cognitive abilities exist, a correct quantification of them may still be pending. These findings suggest that social inequalities in cognitive skills are worrisome, but for a different reason than understood in contemporary debates within social science. A large chunk of the here initially estimated social inequalities in mathematics ability, e.g., against children with a migration background or low-SES children, might be in large part the result of error-prone assessments and of the possible socio-cultural biases present in them. Thus, differences in estimated skills might not be a reflection of inequalities already embodied in children's brains, as supposedly shown by their (in)capacity to solve basic mathematics operations.

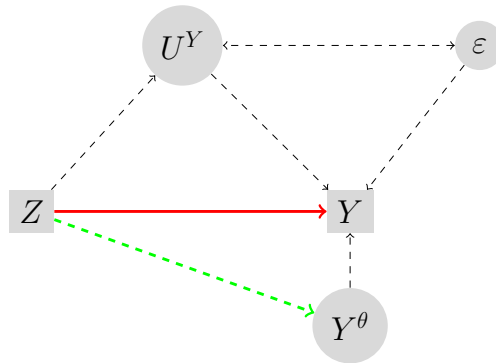
Finally, the findings presented here do lend credence to the replicability crisis in psychology (Loken and Gelman 2017). The replication of measurement operations of psychological constructs might be ill-founded, which probably relates to the fact that, under other scientific conceptions of measurement, psychology's standardized assessments may not have measured anything at all (Trendler 2013). In fact, other studies argued that the ontological basis of cognitive constructs might be more similar to a process happening in the brain than to a quantity residing presumably in the brain (Guyon 2018), a point in which social scientists, in general, may have more to say than current validation practices in psychology. Future

studies should focus on establishing the connections between test scores and measures of brain's structures and functions, which under a causal account of validity seem like a requirement for establishing sound comparisons among children on a "causally" validated metric. And although further requiring that observations obtained through these standardized tests conform to the conditions of a conjoint additive structure, before scaling the data to produce an interval scale, is too strict a criteria for psychological measures to meet, it is the only way one can speak of an interval metric, under a scientific conception of measurement in correspondence with the natural sciences.

## Tables and Figures

Table 8: ACM Checks for Mathematics Test, PPVT-4 and ELFRA-2P

Instrument	Adjacent 3x3 Matrices	
	Weighted Mean	Unweighted Mean
ELFRA-2 Productive subscale	18.1	40.8
PPVT-4	12.5	42.4
Mathematics Test	17.9	22.9



Note: Each node represents an analytical random variable:  $Y$  is the observed or manifest test score;  $Y^\theta$  the unobservable cognitive trait;  $U^Y$  the measurement error of the test;  $Z$  the independent variable of interest; and  $\epsilon$  an error term as would be observed from  $Y = Z\beta + \epsilon$ . Arrows in turn represent causal associations between these variables, with dashed lines for hypothesized but not estimated ones.

Figure 15: Directed Acyclic Graph Illustrating Measurement Error Caused by Measurement Invariance

Table 9: Mokken Scale Analysis Results for Mathematics Test Items

Statistic	Mathematics Test	
	Complete: 20 items	Subscale: 11 items
Number of Unscalable Items at $c=0.3$	7	0
Number of Scales	2	1
Scalability Index H	0.236	0.378
Number of negative item-scale scalability	0	0
Number of negative inter-item scalability	4	0
Monotonicity Violations	1	0
Number of Flagged Items W 1 Index	3	0
Number of Flagged Items W 3 Index	6	0
H_T	0.47	0.539

Table 10: Number and Proportion of Items with DIF by Generalized Mantel-Haezel Test

	Preterm		Girls		Mig. Background		Low-SES	
	j	%	j	%	j	%	j	%
ELFRA-2 Productive subscale	14	5.38	67	25.8	92	35.4	78	30.00
PPVT-4	6	3.66	38	23.2	27	16.5	16	9.76
Mathematics Test	5	25.00	7	35.0	6	30.0	3	15.00

Table 11: Effect Size Estimates from Linear Models in Panel A

Dependent Variable	Covariates	SSR	d.f.	Eta	Delta
PIRT WLE	Preterm	0.26	1	0.0003	0.0002
	Gender	5.58	1	0.0058	0.0047
	Migration Background	0	1	1.71e-06	1.38e-06
	Socioeconomic Status	33.63	3	0.0337	0.0281
	ELFRA-2P	27.82	1	0.0280	0.0233
20 items Sum Score	PPVT-4	71.03	1	0.0686	0.0594
	Preterm	0.45	1	0.0005	0.0004
	Gender	2.91	1	0.0031	0.0024
	Migration Background	0.24	1	0.0003	0.0002
	Socioeconomic Status	42.59	3	0.0435	0.0356
Mokken Sum Score	ELFRA-2P	31.61	1	0.0327	0.0264
	PPVT-4	79.13	1	0.0779	0.0662
	Preterm	1.65	1	0.0017	0.0014
	Gender	4.97	1	0.0051	0.0042
	Migration Background	1.18	1	0.0012	0.0010
	Socioeconomic Status	32.72	3	0.0325	0.0274
	ELFRA-2P	28.45	1	0.0284	0.0238
	PPVT-4	69.12	1	0.0663	0.0578

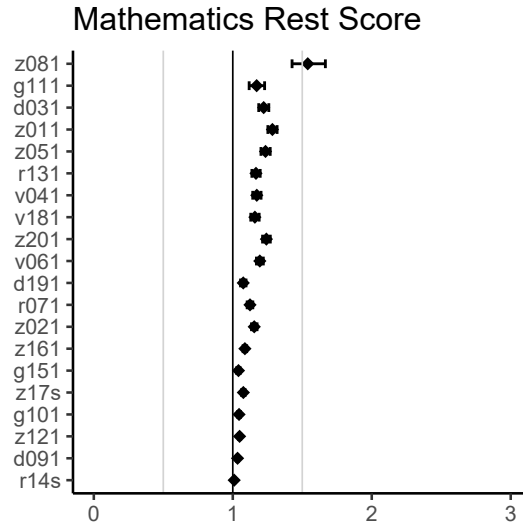


Figure 16: Mathematics Rest-score Relative Risk Adjusted by Rest Score

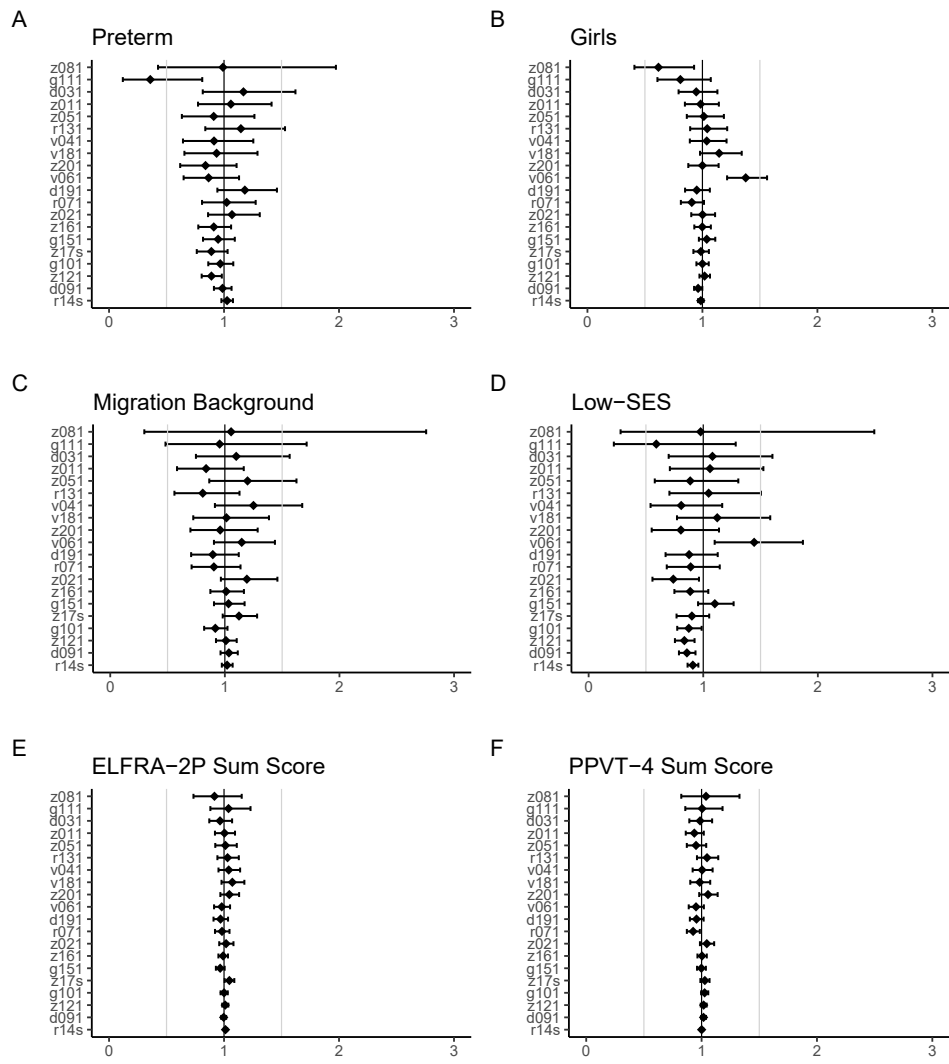


Figure 17: Relative risks for Various Covariates Adjusted by Rest Score

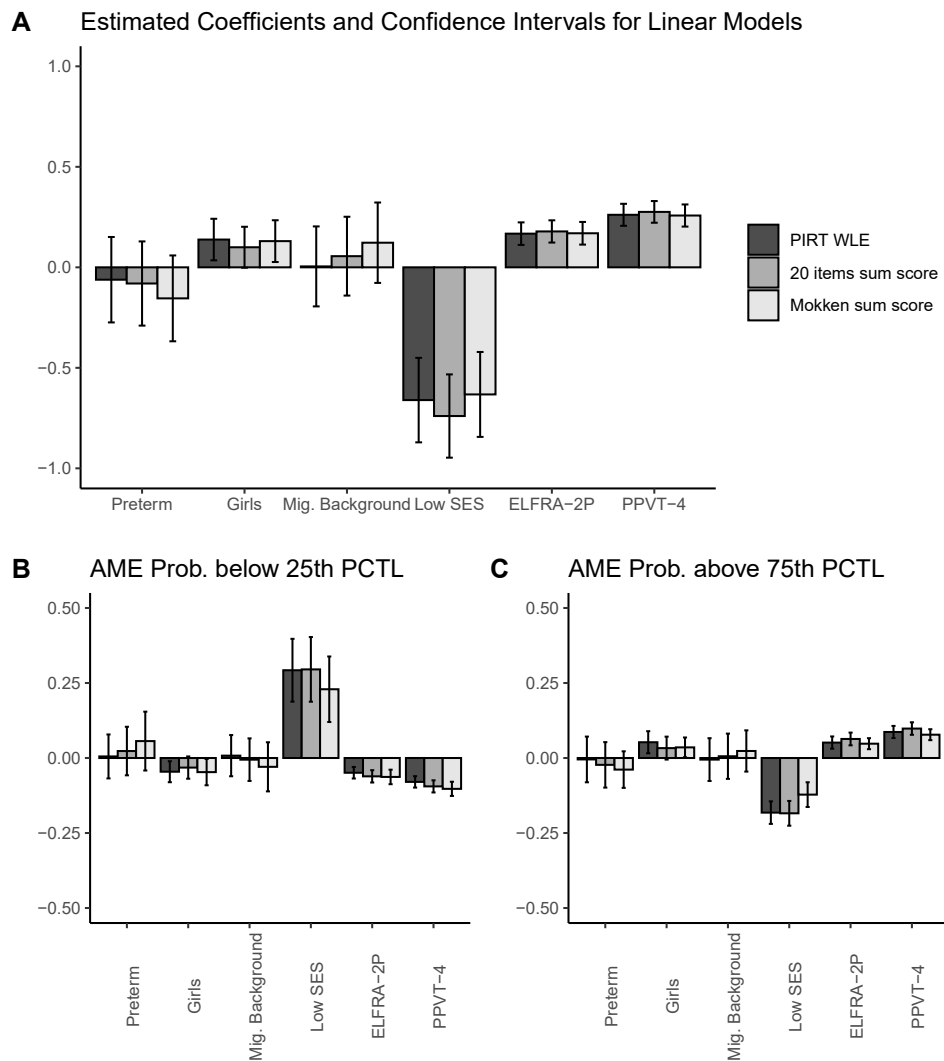


Figure 18: Standardized Estimated Coefficients in Linear Model (Panel A) and Average Marginal Effects from Cumulative Link Model (Panels B and C) for Scores on Mathematics Test

## CONCLUSION

The four empirical chapters presented in this dissertation make the case that furthering causal inference thinking in family sociology, social inequality, social mobility, and family demography research is a worthy endeavor with important implications. Thanks to this causal inference lens, I was able to find little support for some of the most often touted hypotheses in family research – e.g., the negative effects of divorce and family instability, as well as evidence for the causal mediation role of parenting, on children’s wellbeing, and I showed how the measurement of unobservable constructs related to children’s cognitive development is far more complex than acknowledged because these constructs are endogenous to the sociological processes that determine them. Hence, the strong methodological focus of this dissertation opens up the path for further theorizing. One important avenue in theory development is the recognition that the more complex associations between family behavior and children’s wellbeing often assumes models that may lack an appropriate representation in empirical studies. The match, or mismatch, between a given empirical design and a statistical model and a theory, such as those falling under the life course umbrella (L. Bernardi, Huinink, and Settersten Jr 2019), should be evaluated before considering whether any theory is supported by empirical data or not. Staying at the level of associations is not enough to uncover fundamentally causal mechanisms that are assumed to be at play in the reproduction of social inequalities.

The results of Chapter 2 suggest that the departure of the biological father out of the family unit acts as a marker of life course socioeconomic disadvantage, and not as a cause of negative effects on children’s wellbeing, employing data from the Fragile Families and Child Wellbeing Study. Why is it that divorce has no causal effects on children’s wellbeing? This is an

important question that invites us to open up the black box of family events and experiences, and to investigate which reinforcing and counteracting processes are triggered by family changes, so that, despite the hypothesized negative effects, following the perspectives of the family investment and family stress models, we do not find on average a negative effect of father absence on children's wellbeing. However, results related to the socio-emotional dimension of adolescents' wellbeing seem more consistent with small negative effects of this particular family transition. Adolescence, an important life course stage in human development, deserves further scrutiny and specialized studies to understand why father absence effects are more likely to be found in this stage as opposed to experiences of father absence at younger ages.

Chapter 3 explored the role ascribed to family instability, operationalized as the number of family transitions experienced during the first 15 years of a child's life, on the internalizing and externalizing child behavior problem dimension. Employing the same data from the Fragile Families and Child Wellbeing Study, but taking a more dynamic perspective and assuming that previous family changes cause future family changes, I show that treatment-confounder feedback bias affects the estimation of the effects of family instability. Again, reinforcing or counteracting mechanisms are probably at play in selecting families to experience a given number of family transitions. Future research should focus on what those processes are, given that they could hold the promise of becoming policy targets to reduce other potentially negative effects of family instability. However, the results do not support the family instability hypothesis. The effects of various operationalization choices of family instability, capturing early, mid-childhood, or adolescence experiences of family instability, suggest these family transitions are not detrimental to the socio-emotional development of children, independent of the child's gender and racial-ethnic group.

Chapter 4, on the parenting mediation of SES differences in children's language skills, I showed that the mediated share is rather small, small for the large differences seen between the very low SES families and the high SES families in Germany, employing data from a cohort of newborns in the National Educational Panel Study. This suggests that interventions geared at equalizing parenting differences among parents from different SES backgrounds will not be sufficient to make this gap vanish, their impact is rather narrow and limited. These results also show that differences in children's language skills are also informed by experiences that are related to alternative causal pathways, not directly involved in parenting. Those other pathways, in which neighborhoods, peer effects, and larger socialization contexts likely play a predominant role, should be the subject of further research. Collecting data on all of these factors is, however, a difficult challenge to meet. Importantly, causal inference thinking can help in the design of the type of studies and data that could approach such questions in the future.

Finally, chapter 5 took on the topic of measurement error, an often-ignored subject in much of sociological literature. However, measurement error in unobservable constructs can have substantial impacts on the effects we are interested in understanding, particularly on the quantification of social inequalities. The results suggest that the scales used to measure children's skills, based on standardized assessments, applied here to test early numeracy and mathematics skills, do not correspond to the interval scale properties often assumed to exist in continuous scales of cognitive ability that are used to compare children. More important though, I find that measurement error correlates with other social background characteristics, such as low SES family background, which generates endogeneity bias in the estimation of effects of other characteristics we may be interested in.

In this dissertation, I have dealt with important concepts for the wellbeing

of children: parental SES, parenting, family instability, and one specific family transition caused by parental divorce or separation. However, my focus was on the different processes affecting and being affected by these various concepts. A life-course perspective reveals that family instability and parenting should be considered as dynamic concepts that vary over time, rather than static ones, which further makes research on them more complex. Doing so, however, advances debates within family sociology in the direction proposed by Seltzer (2019), by providing a more accurate portrait of the complex, contemporary dynamics of family life.

Another approach in that direction could be to join together anthropological studies of families in contemporary US and Germany with traditional demographic definitions of a household, in conjunction with a life-course perspective that considers the sequence of multiple contexts that are affected by family instability and parenting. For example, as discussed in Bourdieu (1993), it is important to take the ethnomethodological step of suspending the researcher's beliefs in the categories dictated by the family discourse. Behind this discourse lie hidden several presuppositions whose origin remains rather obscure: family as a decision unit, family as private (i.e. concerning only their immediate members' interactions), and as a durable or a stable unit in time (e.g. the family name and the family home). Related to this, future research ought to be aware of the role of family privilege in family theory and research as suggested by Hadfield, Ungar, and Nixon (2018). In future research, we should avoid introducing biases arising from presuppositions on what is good or better on such normative loaded topics, such as the effects of family structure, family instability, or parenting. Achieving this, however, is challenging, because we, as researchers, are also products of our socialization practices. Further reflection on this topic deserves much more scrutiny. In any case, conceiving stable families, specific types of parenting, especially those of families conformed by the two

biological parents of the child, as the benchmark to which other family arrangements ought to be compared to, reminds us of the ethnocentric perspective of mostly White middle-class researchers who insist that children “fare better” under —demographically speaking— living arrangements most similar to their own; while unconvincingly showing biased estimates of effects of family behavior in support of such claims.

In relation to this, the focus of Moynihan (1965) on the ‘sexual mores’ of non-traditional family structures, as well as the early focus on the sexuality of the lower classes at the very onset of demographic research on the family, serves as an important reminder that family studies pass through the studies of sexuality. The connection between reproduction and family formation is better exemplified by studies on contraceptive use in developing countries and teenage pregnancy. The control of sexuality appears, again and again, as the hidden, real target behind some of the social policies targeted at the promotion of stable marriages (Cahill 2005). But, as shown in this paper for family stability, such discussions are much more about what is considered socially and culturally appropriate, and not much about what effects certain family behaviors have on children. Future studies should establish, again from a causal inference perspective, whether a causal effects narrative is supported by the data or not.

By supposedly merely describing how family structure relates to measures of children’s well-being, the effects that configure a particular family formation remain opaque; when in fact they are, at least partially, the result of interlocking trajectories of reinforcing and counteracting mechanisms. If family as a social category contributes to the reproduction of social inequalities, it does so only to the extent that individuals perceive the social world through this particular social category of the nuclear family; and engage in actions and strategies that seek to reproduce nuclear families through institutional sanctioned rituals (Bourdieu 1993). These rituals

guarantee the transmission of different forms of capital – economic, social, and cultural – within one’s own imagined kin; while keeping away from such ideals other individuals who by law (e.g. as was the case of same-sex marriage) or by their socioeconomic circumstances (see Edin and Kissane 2010; and Lauster 2010) lack the means to keep up with the expectations of the ideal family life.

Furthermore, sociological understandings of family dynamics should continue to theorize and empirically assess the role played by the state in the configuration of family structures and their different living standards, as well as the role of parenting in children’s wellbeing (e.g., Maldonado and Nieuwenhuis 2015; Pilkauskas and Micheltore 2019). For example, housing/dwelling urban planning policies to develop suburbs or gated communities for middle class families contrast with the construction of large urban housing projects for low-income families. Both of these policies contribute to the creation of social spaces of different kinds and in which diverging family dynamics are likely to unfold over time, as they are affected by residential context effects (see Sharkey and Faber 2014). Moreover, tax and social welfare policies that benefit or harm individuals in certain social positions also affect the configuration of family forms they choose to live under (Edin and Kissane 2010).

Instead of the disproportionate weight given to psychological explanations of family dynamics, research on the effects of family instability should refocus on understanding how raising children in stable marriage-based two-parent households is dependent on the different social contexts in which families are formed. Variation across countries in the social gradient of the two-parent family shows that similar objective conditions do not necessarily translate into the adoption of the same family structures. However, families are fundamentally the product of the social conditions that enable or make difficult certain family formations, and more research should be geared

towards understanding how the larger social context, such as the social networks in which parents are embedded, structural changes in employment, and family policies more broadly, etc. (Ruggles 2015), can influence the family formation and dissolution dynamics. Some individuals decide to comply with prevalent notions of the ideal family life when they can do so and want to. Others, whose socioeconomic conditions prevent them from abiding by this discourse, seek to build families and raise children in alternative ways nonetheless (Nieuwenhuis and Maldonado 2018); families that suit their contexts amid the adversities which they are faced with.

In this dissertation, I have focused on the topic of family behavior as it relates to children's wellbeing. However, it becomes clear that the modeling and measurement issues here described, and partially dealt with, are far more common, and can be observed in other areas of social inequality research. Family behavior and children's wellbeing were the case studies selected to showcase these data and causal inference problems in an applied empirical case. Future studies should investigate how other claims or stylized facts about social inequality can be the result of these and other forms in which endogeneity biases the results of empirical analyses (Hirschman 2016). There is a broad range of theoretical claims on social inequality that deserve closer scrutiny to advance sociological research and public policies in the causal inference era.

## **Policy implications**

Although the chapters of this dissertation were not focused on policy evaluations, a few implications from some of the findings follow. For the US context, given the little causal evidence for an effect of family instability, policy makers' goals should be the improvement of the socioeconomic circumstances of poor families and not the stabilization of the two-parent family norm. As discussed in chapters 2 and 3, family formation and

dissolution processes are the results of an endogenous decision-making process. In general, individuals will tend to make the family choices they know are more appropriate for their wellbeing and that of their children. Policymakers lack key information about the determinants of those family choices, which should probably be enough reason not to engage in the promotion of a specific type of family life.

Parenting is often touted as a potential great equalizer, fundamental to the reduction of social inequalities in early childhood. Chapter 4 has shown, however, that, although parenting has a positive effect on children's language development, the gaps by parental socioeconomic status are considerably much larger than the largest effect of a hypothetical parenting intervention that would equalize the parenting done by low-SES and high-SES parents. Therefore, parenting may not be the key to the reduction of social inequalities in early childhood, and its role might be quite limited. Inequalities in language skills among children are generated and reproduce through multiple, alternative mechanisms – remaining unobserved – that potentially do not involve parenting. The growing focus on parenting that is seen in the German context – and which is far larger in the US – should be balanced by an equal if not stronger focus on other neglected and contextual (i.e., non-individual or psychological) factors that also shape children's language development in important ways.

## **Limitations**

Despite the advances made in this dissertation, some limitations of the work presented here ought to be acknowledged. First, the focus of each chapter on a single piece of the causal inference puzzle should not overshadow the fact that a joint assessment of these different problems is often necessary in any given empirical application. A unifying framework to evaluate, for example, the mediating role of parenting in explaining the effects of number

of family changes or transitions on an child's wellbeing outcome measured with error would require attention to all the issues raised in these chapters, simultaneously. This work is left for future studies on family dynamics.

Second, family instability can be measured by considering the number of family changes or transitions, as I did in Chapter 3, but is equally important to distinguish different family structures and specific family transitions. These may also matter for child wellbeing. Lack of data containing enough samples of children experiencing very unique parental family life course trajectories prevented me from further exploring this topic - though to some extent the study of father absence, one specific transition, is an example of a study that could be replicated for other very specific, single time point family transitions. The use of register data, in which all childhood family structures experienced by children from birth until the age they leave the parental household, could facilitate the estimation of the effects of specific family structure trajectories. One could compare, for example, children raised in two-parent families with any children experiencing any other family trajectory, on the assumption that a large enough number of cases following unique family trajectories are present in the data, and that enough time-dependent confounders are captured by the register data. Lack of information on important confounders is a serious trade-off to consider when estimating effects of time-varying life course family behaviors employing longitudinal surveys versus register data.

Third, causal mediation is a particularly challenging type of analysis that has not received the attention it truly deserves in the literature. Mediation analyses can, at least in principle, get us closer to the policy relevant processes that transform a given socio-demographic exposure of interest, such as low socioeconomic status, into undesirable social outcomes, lower language skills in children. However, capturing the multiple mediating mechanisms that go from the exposure of interest to the outcome of

relevance, and simultaneously overcoming time-dependent confounding, and in particular exposure-induced confounders of the mediator and outcome effects – which result from multiple and often unintended consequences of such kinds of exposures – can only be partially alleviated by better and more data. To better understand causal mediation, we would need especially designed studies that capture all known or hypothesized mediating effects of parental socioeconomic status effects on children’s outcomes. Causal mediation analysis can make researchers aware of the difficulty of making a mediation claim without warrants on the underlying data generating processes.

Finally, fourth, I have discussed the problem of measurement error in psychological constructs related to children’s cognitive development. However, these errors are far more extended than the realm of children’s wellbeing. A solution to the problem of measurement error was out of the scope of the dissertation, but it remains a crucial challenge to be able to distinguish the signal from the noise in data obtained from the application of standardized assessments for the measurement of cognitive or socio-emotional constructs. Furthermore, measurement error affects many other types of variables – and not just psychological constructs – and the biases introduced by measurement error on outcomes, exposures or mediators require further investigation.

## **Discussion**

Although in order to understand social problems, such as social inequality through intergenerational effects, one must study and understand family dynamics (Smock and Schwartz 2020), social inequality also affects the formation and dissolution of families and, therefore, it is equally important to understand how social inequality can or cannot produce certain effects on the family, and what its consequences might be. Two well known theories

deserve a special comment in light of the results obtained in this dissertation. The results on family instability and the phenomenon of father absence, which I have shown are mostly a marker of socioeconomic disadvantage and not a cause of negative effects, echoes former criticism of the second demographic transition in the direction that the changes in marriage, non-marital cohabitation, divorce, and childbearing outside marriage were the result of a “pattern of disadvantage” and not an attitudinal change or rejection of well-established family norms (Lesthaeghe 2014). If family dynamics do not reflect cultural shifts but changes in socioeconomic circumstances, more attention should be given to understanding the effects of changes to come: automation, home office, reduction in welfare spending, the liberalization and precarization of the labor market, growing wealth inequality with housing being a major driving component of it, climate change, etc., to mention a few of the future challenges for family scholars. It is perhaps those changes that deserve much more attention in light of the results here obtained.

In connection with this, the diverging destinies hypothesis ought to be revisited once more (McLanahan 2004). Although the evidence overwhelmingly suggests that there is a social gradient in marriage and in who follows the normative family formation process across many Western countries (McLanahan and Jacobsen 2015), the effects of such changes in family life on children’s wellbeing are not found. Therefore, it does not follow that such divergence in family life courses is necessarily harmful to children. This is because the two-parent household family structure is only associated with, and not a cause of, children’s wellbeing. As I have shown in this dissertation, children’s wellbeing is a complex function of many time-varying socioeconomic parental characteristics that act in consonance with family life course. When the effects of those time-dependent confounding socioeconomic factors are adjusted for, the diverging family life

courses do not seem to matter as much. In fact, the results of the cumulative instability model suggest that the story behind the family instability hypothesis may have to be entirely reworked given that, as shown in Chapter 5, a higher number of family changes – four or five, a much stronger marker of family instability – are less associated with adolescent problem behavior than a smaller number of changes.

Related to this, previous research on the effects of family instability on children has not fully addressed the reflection problem in the identification of causal effects, a problem which remains in all observational studies – including the ones presented in this dissertation. Family behavior, as shown in the previous chapters, and as past research had discussed (Manski 1993), is a prime example of endogenous behavior (Ginther and Pollak 2004) [pp. 691–693], meaning that family behavior is the result of many other factors informing those behaviors. More awareness of this fact may turn out to be crucial for how researchers study contemporary family dynamics empirically and for further theory development. For example, regardless of how much previous empirical evidence there is on the diverging destinies hypothesis (McLanahan 2004), the stylized fact on which the family instability hypothesis is based will remain ill-founded (Hirschman 2016) until an appropriate identification strategy is found.

However, if we fail to find such effects when improving our methodological tools, then more questions arise. If changes in family behavior do not explain child development, then what does explain it? Why is it that despite families breaking apart are children not harmed? What coping mechanisms are at play in both parents and children? How do these mechanisms inform family formation and dissolution dynamics? Future studies should address these key questions and at the same time try to address the different methodological challenges highlighted in this dissertation: time-varying exposures, time-dependent confounding, causal mediation, and measurement

error. Focusing on the multiple and inter-temporal determinants of individuals' family relationships and parenting behavior could further our understanding of family dynamics and its effects on children (Seltzer 2019). One specific application of this could be the study of whether parenting acts as a mediating mechanism between family instability and children's outcomes as one potential resilience and counteracting mechanism that prevents most children experiencing family instability from obtaining worse outcomes than children raised in the two-parent stable families.

# APPENDIX

## Supplementary Materials - Chapter 2

### Part 1 - Bayesian Additive Regression Trees

In this section of supplementary materials, I provide a more detailed description of Bayesian Additive Regression Trees (BART). As explained in J. Hill and Su (2013), and given the assumption of ignorability conditioning on  $\bar{\mathbf{X}}$ , estimation of the conditional average treatment effect is equivalent to the evaluation of two response surfaces, namely

$$\mathbb{E}(Y(1)|\bar{\mathbf{X}}) = \mathbb{E}(Y|\bar{\mathbf{X}} = \bar{\mathbf{x}}, Z = 1) = f(1, \bar{\mathbf{x}}) \text{ and}$$

$$\mathbb{E}(Y(0)|\bar{\mathbf{X}}) = \mathbb{E}(Y|\bar{\mathbf{X}} = \bar{\mathbf{x}}, Z = 0) = f(0, \bar{\mathbf{x}});$$

one for children who experienced father absence up to a certain time point and another one for children who remained in stable families. BART can flexibly estimate  $f$  in a nonparametric fashion, accounting for nonlinear main effects (e.g., potential quadratic effects of income or age of parents), multiple-way interactions (e.g., between unemployment and housing stability, or alcohol problems and father-mother relationship quality), as well as a high number of covariates (Chipman et al. 2010).

The BART algorithm consists of two main elements: a sum-of-trees and a regularization prior. A binary tree structure  $T$  is a set of sequential decision rules based on the information of the confounder covariates  $\bar{\mathbf{X}}$ , a rule that partitions the confounder covariate space (J. Hill, Linero, and Murray 2020). For example, when trying to predict the event of departure of the biological father, given that this event might be more prevalent among people below the median wage, and in that group, the probability might be higher among people age below 25, and further down the tree, the probability might be higher for low educated Black parents than for other racial-ethnic groups, etc. Assuming a tree  $T$  with  $b$  bottom terminal nodes or leaf functions, for each of these nodes there will be an associated parameter  $\mu_k$ , such that the

set of final nodes of the tree can be denoted by  $M = \{\mu_1, \mu_2, \dots, \mu_b\}$ . A value for the conditional expectation is assigned to each observation unit given the observed data by following the branches of the tree leading to a particular terminal node at the bottom of the tree. The sum-of-trees element is thus a sum of  $J$  of these binary trees which corresponds to a model for the outcome variable  $Y$  such that multiple trees are combined. For each  $(z, \bar{\mathbf{x}})$  data pair, BART follows specific branches of multiple trees, then assigns the value of each terminal node, and adds them all up together. Thus, each output value, denoted by  $g(z, \bar{\mathbf{x}}; T_j, M_j)$ , is obtained from each tree and is then additively combined to provide an estimate of the response surface as in

$$Y = g(z, \bar{\mathbf{x}}; T_1, M_1) + g(z, \bar{\mathbf{x}}; T_2, M_2) + \dots + g(z, \bar{\mathbf{x}}; T_m, M_m) + \epsilon = \sum_{j=1}^J g(z, \bar{\mathbf{x}}; T_j, M_j) + \epsilon \quad (3)$$

where  $T_j$  is the  $j^{th}$  binary tree structure and  $M_j = \{\mu_{j1}, \mu_{j2}, \dots, \mu_{jb}\}$  contains the terminal node parameters which are associated with the  $T_j$  tree.

The second element of BART refers to regularization priors  $p[(T_1, M_1), \dots, (T_m, M_m), \sigma]$ : the number of trees, the variables on which to split, as well as their values, and other parameters. Priors prevent BART from overfitting the model to the data by specifying the size of each of the trees, the shrinkage used to fit each tree, and the degrees of freedom corresponding to the residual standard error (Chipman et al. 2010; and J. Hill 2011), such that the salience, complexity, and size of each tree is reduced in the final model; making each tree a weak learner (J. Hill, Linero, and Murray 2020). To arrive at a prior distribution, these priors are all assumed to be independent of each other and were chosen following the description of J. Hill (2011). BART's algorithm is based on Markov chain Monte Carlo methods, similarly to how ensemble learning works for boosting

(Friedman 2002), and all parameters are estimated in a unifying way as features of the posterior predictive distribution of  $Y$ . With each iteration of the Markov chain, a draw of  $f$  is taken out of the posterior predictive distribution. By denoting  $f^r$  as the  $r^{th}$  draw from this distribution (i.e., a predicted value for children experiencing the departure of the biological father or a stable family structure), it is possible to compute the difference  $d_i^r = f^r(1, \bar{\mathbf{x}}_i) - f^r(0, \bar{\mathbf{x}}_i)$  for each  $i = 1, \dots, n$  and each draw  $r$ . Thus, the average of  $d_i^r$  over  $n$ , with  $r$  fixed, is an approximation to the posterior distribution of the conditional average treatment effect (CATE) in the sample and an estimation of the treatment effect of interest.

To tackle the potential lack of common support, a more robust version of BART can be estimated. The notion of common *causal* support, as advanced by J. Hill and Su (2013), suggests that not all confounders may be equally strong or important and that therefore not all units ought to be included in the estimation of the causal effects (e.g., avoiding to match on variables that don't act as true confounders). In principle, the computation of the effect of father absence should only consider units with a non-zero probability of receiving the treatment across the covariate space. Therefore, some of the observations which lack common support, or which have high posterior uncertainty, could be excluded from the prediction of the counterfactual outcomes. For a more detailed exposition of the specific aspects of the implementation of BART, see J. Hill (2011) and J. Hill, Linero, and Murray (2020); and for a tutorial on how to perform BART in biostatistics that can be easily transported to research on demography, see Tan and Roy (2019).

## **Part 2 - Additional Tables with descriptive statistics**

Table S1: Descriptive Statistics for Analytical Variables in the FFCWS at Baseline by Type of Transition at each wave (baseline, 2nd and 3rd waves)

	Full Sample n (%)	Stable at W2	Unstable at W2	Stable at W3	Unstable at W3
<b>Baby's gender assigned at birth</b>					
Boy	1,041 (52)	891 (54)	150 (44)	724 (54)	134 (57)
Girl	956 (48)	763 (46)	193 (56)	629 (46)	101 (43)
<b>Low birth weight</b>					
Yes	165 (8)	129 (8)	36 (10)	98 (7)	27 (11)
No	1,832 (92)	1,525 (92)	307 (90)	1,255 (93)	208 (89)
<b>Mother's age</b>					
median	25	26	23	27	23
mean (sd)	26.22 (6.07)	26.70 (6.14)	23.89 (5.14)	27.21 (6.14)	24.66 (5.91)
<b>Mother's education</b>					
Less than High-school	587 (29)	453 (27)	134 (39)	342 (25)	85 (36)
High-school or equivalent	548 (27)	429 (26)	119 (35)	334 (25)	68 (29)
Some college, technical education	511 (26)	434 (26)	77 (22)	358 (26)	64 (27)
College or graduate degree	351 (18)	338 (20)	13 (4)	319 (24)	18 (8)
<b>Mother's race-ethnicity</b>					
Other	99 (5)	89 (5)	10 (3)	82 (6)	6 (3)
Hispanic	658 (33)	563 (34)	95 (28)	479 (35)	69 (29)
Black, non-hispanic	597 (30)	430 (26)	167 (49)	297 (22)	103 (44)
White, non-hispanic	643 (32)	572 (35)	71 (21)	495 (37)	57 (24)
<b>Mother's immigration status</b>					
Immigrant	411 (21)	378 (23)	33 (10)	333 (25)	38 (16)
Native	1,586 (79)	1,276 (77)	310 (90)	1,020 (75)	197 (84)
<b>Mother lived with both parents as teenager</b>					
Yes	1,016 (51)	891 (54)	125 (36)	758 (56)	107 (46)
No	981 (49)	763 (46)	218 (64)	595 (44)	128 (54)
<b>Mother's religiosity</b>					
At least once a year	1,223 (61)	1,046 (63)	177 (52)	887 (66)	127 (54)
Hardly, never	774 (39)	608 (37)	166 (48)	466 (34)	108 (46)

Table S1: Descriptive Statistics for Analytical Variables in the FFCWS at Baseline by Type of Transition at each wave (baseline, 2nd and 3rd waves) (*continued*)

	Full Sample n (%)	Stable at W2	Unstable at W2	Stable at W3	Unstable at W3
<b>Mother was suggested/thought about abortion</b>					
No	1,604 (80)	1,364 (82)	240 (70)	1,148 (85)	174 (74)
Yes	393 (20)	290 (18)	103 (30)	205 (15)	61 (26)
<b>Mother drank alcohol during pregnancy</b>					
Never	1,800 (90)	1,489 (90)	311 (91)	1,218 (90)	215 (91)
Less than one time per month	163 (8)	137 (8)	26 (8)	114 (8)	16 (7)
Several times per month	22 (1)	18 (1)	4 (1)	13 (1)	2 (1)
Several times per week	10 (1)	9 (1)	1 (0)	7 (1)	2 (1)
Every day	2 (0)	1 (0)	1 (0)	1 (0)	0 (0)
<b>Mother smoked during pregnancy</b>					
None	1,661 (83)	1,406 (85)	255 (74)	1,169 (86)	189 (80)
< 1 pck./day	291 (15)	217 (13)	74 (22)	161 (12)	41 (17)
Btw. < 1 and 2 pck./day	42 (2)	29 (2)	13 (4)	21 (2)	5 (2)
> 2 pck./day	3 (0)	2 (0)	1 (0)	2 (0)	0 (0)
<b>Mother took drugs during pregnancy</b>					
Never	1,938 (97)	1,618 (98)	320 (93)	1,327 (98)	228 (97)
Less than 1 time per month	36 (2)	24 (1)	12 (3)	19 (1)	4 (2)
Several times per month	13 (1)	7 (0)	6 (2)	4 (0)	2 (1)
Several times per week	5 (0)	2 (0)	3 (1)	1 (0)	0 (0)
Every day	5 (0)	3 (0)	2 (1)	2 (0)	1 (0)
<b>Mother's score on PPVT test at wave 3</b>					
median	91	92	88	93	88
mean (sd)	91.86 (12.79)	92.61 (13.08)	88.24 (10.61)	93.59 (13.11)	88.05 (12.41)
<b>Father's age</b>					
median	28	29	25	29	26
mean (sd)	28.80 (6.96)	29.32 (7.00)	26.29 (6.20)	29.69 (6.86)	27.59 (7.23)
<b>Father's education</b>					
Less than High-school	617 (31)	477 (29)	140 (41)	365 (27)	87 (37)
High-school or equivalent	566 (28)	447 (27)	119 (35)	345 (25)	83 (35)

Table S1: Descriptive Statistics for Analytical Variables in the FFCWS at Baseline by Type of Transition at each wave (baseline, 2nd and 3rd waves) (*continued*)

	Full Sample n (%)	Stable at W2	Unstable at W2	Stable at W3	Unstable at W3
Some college, technical education	491 (25)	420 (25)	71 (21)	349 (26)	52 (22)
College or graduate degree	323 (16)	310 (19)	13 (4)	294 (22)	13 (6)
<b>Father's race-ethnicity</b>					
Other	107 (5)	96 (6)	11 (3)	82 (6)	11 (5)
Hispanic	672 (34)	556 (34)	116 (34)	464 (34)	74 (31)
Black, non-hispanic	625 (31)	463 (28)	162 (47)	330 (24)	104 (44)
White, non-hispanic	593 (30)	539 (33)	54 (16)	477 (35)	46 (20)
<b>Father's immigration status</b>					
Immigrant	428 (21)	387 (23)	41 (12)	341 (25)	40 (17)
Native	1,569 (79)	1,267 (77)	302 (88)	1,012 (75)	195 (83)
<b>Father lived with both parents as teenager</b>					
Yes	1,066 (53)	940 (57)	126 (37)	812 (60)	102 (43)
No	931 (47)	714 (43)	217 (63)	541 (40)	133 (57)
<b>Father's religiosity</b>					
At least once a year	1,122 (56)	964 (58)	158 (46)	808 (60)	123 (52)
Hardly, never	875 (44)	690 (42)	185 (54)	545 (40)	112 (48)
<b>Father suggested/thought about abortion</b>					
Yes	280 (14)	200 (12)	80 (23)	145 (11)	41 (17)
No	1,717 (86)	1,454 (88)	263 (77)	1,208 (89)	194 (83)
<b>Father's last name on birth certificate</b>					
Yes	1,931 (97)	1,608 (97)	323 (94)	1,329 (98)	216 (92)
No	66 (3)	46 (3)	20 (6)	24 (2)	19 (8)
<b>Relatives/non-relatives in household (1st wave)</b>					
Yes	566 (28)	437 (26)	129 (38)	339 (25)	76 (32)
No	1,431 (72)	1,217 (74)	214 (62)	1,014 (75)	159 (68)
<b>Child lives with siblings from mother side (1st wave)</b>					
Yes	1,201 (60)	996 (60)	205 (60)	816 (60)	138 (59)
No	796 (40)	658 (40)	138 (40)	537 (40)	97 (41)

Table S1: Descriptive Statistics for Analytical Variables in the FFCWS at Baseline by Type of Transition at each wave (baseline, 2nd and 3rd waves) (*continued*)

	Full Sample n (%)	Stable at W2	Unstable at W2	Stable at W3	Unstable at W3
<b>Mother's general health (1st wave)</b>					
Fair or poor	131 (7)	105 (6)	26 (8)	81 (6)	18 (8)
Good	1,866 (93)	1,549 (94)	317 (92)	1,272 (94)	217 (92)
<b>Mother had alcohol problems (1st wave)</b>					
Yes	39 (2)	26 (2)	13 (4)	19 (1)	4 (2)
No	1,958 (98)	1,628 (98)	330 (96)	1,334 (99)	231 (98)
<b>Father had alcohol problems(1st wave)</b>					
Yes	100 (5)	72 (4)	28 (8)	50 (4)	18 (8)
No	1,897 (95)	1,582 (96)	315 (92)	1,303 (96)	217 (92)
<b>Relationship quality between mother and father (after pregnant)</b>					
Same	860 (43)	724 (44)	136 (40)	584 (43)	112 (48)
Worse	114 (6)	82 (5)	32 (9)	60 (4)	15 (6)
Better	1,023 (51)	848 (51)	175 (51)	709 (52)	108 (46)
<b>Violence against mother, e.g., physical and verbal abuse (1st wave)</b>					
Yes	543 (27)	427 (26)	116 (34)	335 (25)	72 (31)
No	1,454 (73)	1,227 (74)	227 (66)	1,018 (75)	163 (69)
<b>Father was in jail (1st wave)</b>					
Yes	28 (1)	13 (1)	15 (4)	9 (1)	2 (1)
No	1,969 (99)	1,641 (99)	328 (96)	1,344 (99)	233 (99)
<b>Public assistance, e.g., TANF or food stamps (1st wave)</b>					
Yes	533 (27)	386 (23)	147 (43)	275 (20)	81 (34)
No	1,464 (73)	1,268 (77)	196 (57)	1,078 (80)	154 (66)
<b>Relatives provided financial assistance (1st wave)</b>					
Yes	719 (36)	537 (32)	182 (53)	394 (29)	102 (43)
No	1,278 (64)	1,117 (68)	161 (47)	959 (71)	133 (57)
<b>Poverty Categories based on Household income (1st wave)</b>					
More than 300%	664 (33)	604 (37)	60 (17)	544 (40)	51 (22)
Btw. 200-299%	342 (17)	290 (18)	52 (15)	237 (18)	39 (17)

Table S1: Descriptive Statistics for Analytical Variables in the FFCWS at Baseline by Type of Transition at each wave (baseline, 2nd and 3rd waves) (*continued*)

	Full Sample n (%)	Stable at W2	Unstable at W2	Stable at W3	Unstable at W3
Btw. 100-199%	475 (24)	373 (23)	102 (30)	295 (22)	63 (27)
Btw. 50-99%	257 (13)	200 (12)	57 (17)	150 (11)	34 (14)
Btw. 0-49%	259 (13)	187 (11)	72 (21)	127 (9)	48 (20)
<b>Equivalized Household income (1st wave)</b>					
min	0	0	0	0	0
max	94575.53	94575.53	89974.39	94575.53	89974.39
mean (sd)	23,601.08 (22,582.39)	25,374.93 (23,562.72)	15,047.29 (14,290.23)	27,310.96 (24,424.96)	17,542.13 (17,911.69)
median	15909.9	17320.51	11250.00	19006.58	12374.37
<b>Father's occupation (1st wave)</b>					
White collar, high skill	325 (16)	300 (18)	25 (7)	268 (20)	24 (10)
Services, high skill	562 (28)	465 (28)	97 (28)	369 (27)	75 (32)
Manual blue collar	696 (35)	572 (35)	124 (36)	475 (35)	76 (32)
Other low skill	194 (10)	160 (10)	34 (10)	122 (9)	30 (13)
Self-employed	16 (1)	14 (1)	2 (1)	13 (1)	1 (0)
Unemployed	138 (7)	94 (6)	44 (13)	66 (5)	21 (9)
OLF	66 (3)	49 (3)	17 (5)	40 (3)	8 (3)
<b>Mother worked before having child</b>					
Yes	1,619 (81)	1,347 (81)	272 (79)	1,102 (81)	191 (81)
No	378 (19)	307 (19)	71 (21)	251 (19)	44 (19)
<b>Did you change residence since previous wave? (1st wave)</b>					
Yes	868 (43)	708 (43)	160 (47)	553 (41)	116 (49)
No	1,129 (57)	946 (57)	183 (53)	800 (59)	119 (51)
<b>Current living situation (1st wave)</b>					
Rent	1,255 (63)	998 (60)	257 (75)	776 (57)	171 (73)
Owned house/apt.	742 (37)	656 (40)	86 (25)	577 (43)	64 (27)
<b>Neighborhood safety (1st wave)</b>					
Very unsafe	40 (2)	29 (2)	11 (3)	21 (2)	7 (3)
Unsafe	212 (11)	158 (10)	54 (16)	123 (9)	30 (13)
Safe	1,117 (56)	920 (56)	197 (57)	736 (54)	142 (60)

Table S1: Descriptive Statistics for Analytical Variables in the FFCWS at Baseline by Type of Transition at each wave (baseline, 2nd and 3rd waves) (*continued*)

	Full Sample n (%)	Stable at W2	Unstable at W2	Stable at W3	Unstable at W3
Very Safe	628 (31)	547 (33)	81 (24)	473 (35)	56 (24)

Table S2: Descriptive Statistics for Analytical Variables in the FFCWS at Baseline by Type of Transition at each wave (4th - 6th waves)

	Stable at W4 n (%)	Unstable at W4	Stable at W5	Unstable at W5	Stable at W6	Unstable at W6
<b>Baby's gender assigned at birth</b>						
Boy	622 (54)	77 (47)	503 (54)	81 (53)	383 (54)	89 (52)
Girl	521 (46)	87 (53)	429 (46)	72 (47)	332 (46)	81 (48)
<b>Low birth weight</b>						
Yes	77 (7)	16 (10)	57 (6)	12 (8)	42 (6)	11 (6)
No	1,066 (93)	148 (90)	875 (94)	141 (92)	673 (94)	159 (94)
<b>Mother's age</b>						
median	28	24	28	24	29	27
mean (sd)	27.59 (6.10)	25.34 (6.12)	28.09 (6.05)	25.72 (6.11)	28.51 (5.95)	27.48 (6.08)
<b>Mother's education</b>						
Less than High-school	268 (23)	60 (37)	200 (21)	49 (32)	145 (20)	36 (21)
High-school or equivalent	279 (24)	41 (25)	213 (23)	45 (29)	148 (21)	51 (30)
Some college, technical education	301 (26)	44 (27)	248 (27)	39 (25)	189 (26)	49 (29)
College or graduate degree	295 (26)	19 (12)	271 (29)	20 (13)	233 (33)	34 (20)
<b>Mother's race-ethnicity</b>						
Other	75 (7)	5 (3)	65 (7)	8 (5)	52 (7)	11 (6)
Hispanic	393 (34)	70 (43)	317 (34)	55 (36)	234 (33)	60 (35)
Black, non-hispanic	234 (20)	56 (34)	177 (19)	42 (27)	134 (19)	37 (22)
White, non-hispanic	441 (39)	33 (20)	373 (40)	48 (31)	295 (41)	62 (36)
<b>Mother's immigration status</b>						
Immigrant	298 (26)	27 (16)	258 (28)	29 (19)	207 (29)	42 (25)
Native	845 (74)	137 (84)	674 (72)	124 (81)	508 (71)	128 (75)
<b>Mother lived with both parents as teenager</b>						
Yes	667 (58)	73 (45)	587 (63)	62 (41)	469 (66)	94 (55)
No	476 (42)	91 (55)	345 (37)	91 (59)	246 (34)	76 (45)
<b>Mother's religiosity</b>						
At least once a year	751 (66)	107 (65)	635 (68)	93 (61)	508 (71)	99 (58)
Hardly, never	392 (34)	57 (35)	297 (32)	60 (39)	207 (29)	71 (42)

Table S2: Descriptive Statistics for Analytical Variables in the FFCWS at Baseline by Type of Transition at each wave (4th - 6th waves) (*continued*)

	Stable at W4 n (%)	Unstable at W4	Stable at W5	Unstable at W5	Stable at W6	Unstable at W6
<b>Mother was suggested/thought about abortion</b>						
No	984 (86)	127 (77)	823 (88)	120 (78)	636 (89)	147 (86)
Yes	159 (14)	37 (23)	109 (12)	33 (22)	79 (11)	23 (14)
<b>Mother drank alcohol during pregnancy</b>						
Never	1,028 (90)	149 (91)	843 (90)	137 (90)	646 (90)	152 (89)
Less than one time per month	100 (9)	10 (6)	77 (8)	14 (9)	59 (8)	16 (9)
Several times per month	10 (1)	2 (1)	9 (1)	1 (1)	8 (1)	1 (1)
Several times per week	4 (0)	3 (2)	3 (0)	1 (1)	2 (0)	1 (1)
Every day	1 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
<b>Mother smoked during pregnancy</b>						
None	1,000 (87)	135 (82)	838 (90)	123 (80)	650 (91)	147 (86)
< 1 pck./day	126 (11)	23 (14)	82 (9)	28 (18)	56 (8)	21 (12)
Btw. < 1 and 2 pck./day	16 (1)	5 (3)	11 (1)	2 (1)	8 (1)	2 (1)
> 2 pck./day	1 (0)	1 (1)	1 (0)	0 (0)	1 (0)	0 (0)
<b>Mother took drugs during pregnancy</b>						
Never	1,122 (98)	159 (97)	921 (99)	149 (97)	708 (99)	166 (98)
Less than 1 time per month	16 (1)	3 (2)	9 (1)	2 (1)	6 (1)	3 (2)
Several times per month	3 (0)	1 (1)	1 (0)	2 (1)	1 (0)	0 (0)
Several times per week	1 (0)	0 (0)	1 (0)	0 (0)	0 (0)	1 (1)
Every day	1 (0)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)
<b>Mother's score on PPVT test at wave 3</b>						
median	93	91	94	92	95	91
mean (sd)	94.02 (13.33)	90.74 (11.96)	94.43 (13.55)	92.38 (12.50)	95.11 (13.68)	92.39 (12.64)
<b>Father's age</b>						
median	30	27	30	27	30	30
mean (sd)	30.05 (6.81)	27.57 (6.65)	30.42 (6.74)	28.76 (7.20)	30.82 (6.75)	29.78 (6.53)
<b>Father's education</b>						
Less than High-school	289 (25)	61 (37)	216 (23)	47 (31)	147 (21)	45 (26)
High-school or equivalent	269 (24)	55 (34)	203 (22)	55 (36)	151 (21)	45 (26)

Table S2: Descriptive Statistics for Analytical Variables in the FFCWS at Baseline by Type of Transition at each wave (4th - 6th waves) (*continued*)

	Stable at W4 n (%)	Unstable at W4	Stable at W5	Unstable at W5	Stable at W6	Unstable at W6
Some college, technical education	308 (27)	34 (21)	254 (27)	38 (25)	199 (28)	46 (27)
College or graduate degree	277 (24)	14 (9)	259 (28)	13 (8)	218 (30)	34 (20)
<b>Father's race-ethnicity</b>						
Other	77 (7)	4 (2)	60 (6)	14 (9)	48 (7)	10 (6)
Hispanic	373 (33)	72 (44)	298 (32)	55 (36)	220 (31)	56 (33)
Black, non-hispanic	263 (23)	59 (36)	200 (21)	46 (30)	148 (21)	44 (26)
White, non-hispanic	430 (38)	29 (18)	374 (40)	38 (25)	299 (42)	60 (35)
<b>Father's immigration status</b>						
Immigrant	300 (26)	33 (20)	252 (27)	33 (22)	200 (28)	39 (23)
Native	843 (74)	131 (80)	680 (73)	120 (78)	515 (72)	131 (77)
<b>Father lived with both parents as teenager</b>						
Yes	702 (61)	80 (49)	602 (65)	73 (48)	470 (66)	101 (59)
No	441 (39)	84 (51)	330 (35)	80 (52)	245 (34)	69 (41)
<b>Father's religiosity</b>						
At least once a year	686 (60)	99 (60)	574 (62)	79 (52)	445 (62)	99 (58)
Hardly, never	457 (40)	65 (40)	358 (38)	74 (48)	270 (38)	71 (42)
<b>Father suggested/thought about abortion</b>						
Yes	112 (10)	27 (16)	88 (9)	17 (11)	67 (9)	15 (9)
No	1,031 (90)	137 (84)	844 (91)	136 (89)	648 (91)	155 (91)
<b>Father's last name on birth certificate</b>						
Yes	1,125 (98)	158 (96)	922 (99)	147 (96)	711 (99)	164 (96)
No	18 (2)	6 (4)	10 (1)	6 (4)	4 (1)	6 (4)
<b>Relatives/non-relatives in household (1st wave)</b>						
Yes	269 (24)	53 (32)	207 (22)	47 (31)	145 (20)	45 (26)
No	874 (76)	111 (68)	725 (78)	106 (69)	570 (80)	125 (74)
<b>Child lives with siblings from mother side (1st wave)</b>						
Yes	689 (60)	101 (62)	554 (59)	100 (65)	418 (58)	104 (61)
No	454 (40)	63 (38)	378 (41)	53 (35)	297 (42)	66 (39)

Table S2: Descriptive Statistics for Analytical Variables in the FFCWS at Baseline by Type of Transition at each wave (4th - 6th waves) (*continued*)

	Stable at W4 n (%)	Unstable at W4	Stable at W5	Unstable at W5	Stable at W6	Unstable at W6
<b>Mother's general health (1st wave)</b>						
Fair or poor	66 (6)	15 (9)	46 (5)	15 (10)	31 (4)	13 (8)
Good	1,077 (94)	149 (91)	886 (95)	138 (90)	684 (96)	157 (92)
<b>Mother had alcohol problems (1st wave)</b>						
Yes	15 (1)	3 (2)	9 (1)	3 (2)	5 (1)	3 (2)
No	1,128 (99)	161 (98)	923 (99)	150 (98)	710 (99)	167 (98)
<b>Father had alcohol problems(1st wave)</b>						
Yes	39 (3)	8 (5)	30 (3)	5 (3)	22 (3)	3 (2)
No	1,104 (97)	156 (95)	902 (97)	148 (97)	693 (97)	167 (98)
<b>Relationship quality between mother and father (after pregnant)</b>						
Same	503 (44)	69 (42)	408 (44)	64 (42)	313 (44)	74 (44)
Worse	43 (4)	14 (9)	32 (3)	9 (6)	24 (3)	7 (4)
Better	597 (52)	81 (49)	492 (53)	80 (52)	378 (53)	89 (52)
<b>Violence against mother, e.g., physical and verbal abuse (1st wave)</b>						
Yes	277 (24)	51 (31)	213 (23)	51 (33)	167 (23)	43 (25)
No	866 (76)	113 (69)	719 (77)	102 (67)	548 (77)	127 (75)
<b>Father was in jail (1st wave)</b>						
Yes	7 (1)	1 (1)	4 (0)	2 (1)	3 (0)	1 (1)
No	1,136 (99)	163 (99)	928 (100)	151 (99)	712 (100)	169 (99)
<b>Public assistance, e.g., TANF or food stamps (1st wave)</b>						
Yes	209 (18)	49 (30)	144 (15)	49 (32)	93 (13)	36 (21)
No	934 (82)	115 (70)	788 (85)	104 (68)	622 (87)	134 (79)
<b>Relatives provided financial assistance (1st wave)</b>						
Yes	304 (27)	67 (41)	233 (25)	53 (35)	167 (23)	45 (26)
No	839 (73)	97 (59)	699 (75)	100 (65)	548 (77)	125 (74)
<b>Poverty Categories based on Household income (1st wave)</b>						
More than 300%	492 (43)	39 (24)	433 (46)	44 (29)	364 (51)	61 (36)
Btw. 200-299%	191 (17)	36 (22)	165 (18)	18 (12)	114 (16)	43 (25)

Table S2: Descriptive Statistics for Analytical Variables in the FFCWS at Baseline by Type of Transition at each wave (4th - 6th waves) (*continued*)

	Stable at W4 n (%)	Unstable at W4	Stable at W5	Unstable at W5	Stable at W6	Unstable at W6
Btw. 100-199%	251 (22)	35 (21)	180 (19)	47 (31)	136 (19)	32 (19)
Btw. 50-99%	117 (10)	30 (18)	92 (10)	21 (14)	65 (9)	20 (12)
Btw. 0-49%	92 (8)	24 (15)	62 (7)	23 (15)	36 (5)	14 (8)
<b>Equivalized Household income (1st wave)</b>						
min	0	0	0	0	0.000	1530.931
max	94575.53	89974.39	94575.53	89974.39	94575.53	94575.53
mean (sd)	28,849.71 (25,086.07)	17,991.71 (16,346.87)	31,037.28 (25,904.28)	19,558.24 (19,016.54)	32,957.43 (26,310.14)	27,123.76 (24,486.69)
median	21213.2	15000.0	21250	12500	24537.39	19006.58
<b>Father's occupation (1st wave)</b>						
White collar, high skill	250 (22)	14 (9)	226 (24)	17 (11)	192 (27)	28 (16)
Services, high skill	309 (27)	43 (26)	250 (27)	45 (29)	186 (26)	54 (32)
Manual blue collar	398 (35)	61 (37)	321 (34)	55 (36)	236 (33)	66 (39)
Other low skill	95 (8)	24 (15)	76 (8)	15 (10)	57 (8)	14 (8)
Self-employed	9 (1)	4 (2)	8 (1)	0 (0)	6 (1)	1 (1)
Unemployed	50 (4)	11 (7)	32 (3)	11 (7)	24 (3)	6 (4)
OLF	32 (3)	7 (4)	19 (2)	10 (7)	14 (2)	1 (1)
<b>Mother worked before having child</b>						
Yes	925 (81)	139 (85)	744 (80)	129 (84)	574 (80)	137 (81)
No	218 (19)	25 (15)	188 (20)	24 (16)	141 (20)	33 (19)
<b>Did you change residence since previous wave? (1st wave)</b>						
Yes	456 (40)	78 (48)	358 (38)	69 (45)	255 (36)	75 (44)
No	687 (60)	86 (52)	574 (62)	84 (55)	460 (64)	95 (56)
<b>Current living situation (1st wave)</b>						
Rent	632 (55)	113 (69)	501 (54)	93 (61)	366 (51)	101 (59)
Owned house/apt.	511 (45)	51 (31)	431 (46)	60 (39)	349 (49)	69 (41)
<b>Neighborhood safety (1st wave)</b>						
Very unsafe	16 (1)	4 (2)	10 (1)	5 (3)	6 (1)	3 (2)
Unsafe	96 (8)	16 (10)	64 (7)	24 (16)	48 (7)	13 (8)
Safe	620 (54)	96 (59)	506 (54)	82 (54)	383 (54)	95 (56)

Table S2: Descriptive Statistics for Analytical Variables in the FFCWS at Baseline by Type of Transition at each wave (4th - 6th waves) (*continued*)

	Stable at W4 n (%)	Unstable at W4	Stable at W5	Unstable at W5	Stable at W6	Unstable at W6
Very Safe	411 (36)	48 (29)	352 (38)	42 (27)	278 (39)	59 (35)

Table S3: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 3

	Stable at W3 n (%)	Unstable at W3
<b>Relatives/non-relatives in household (1st wave)</b>		
Yes	339 (25)	76 (32)
No	1,014 (75)	159 (68)
<b>Relatives/non-relatives in household (2nd wave)</b>		
Yes	254 (19)	51 (22)
No	1,099 (81)	184 (78)
<b>Child lives with siblings from mother side (1st wave)</b>		
Yes	816 (60)	138 (59)
No	537 (40)	97 (41)
<b>Child lives with siblings from mother side (2nd wave)</b>		
Yes	1,239 (92)	186 (79)
No	114 (8)	49 (21)
<b>Multipartner Fertility (2nd wave)</b>		
Yes	347 (26)	74 (31)
No	1,006 (74)	161 (69)
<b>Mother's general health (1st wave)</b>		
Fair or poor	81 (6)	18 (8)
Good	1,272 (94)	217 (92)
<b>Mother's general health (2nd wave)</b>		
Fair or poor	135 (10)	34 (14)
Good	1,218 (90)	201 (86)
<b>Mother meets depression criteria (2nd wave)</b>		
Yes	146 (11)	36 (15)
No	1,207 (89)	199 (85)
<b>Mother had alcohol problems (1st wave)</b>		
Yes	19 (1)	4 (2)
No	1,334 (99)	231 (98)
<b>Mother had alcohol problems (2nd wave)</b>		
Yes	9 (1)	5 (2)
No	1,344 (99)	230 (98)
<b>Father had alcohol problems (1st wave)</b>		
Yes	50 (4)	18 (8)
No	1,303 (96)	217 (92)
<b>Father had alcohol problems (2nd wave)</b>		
Yes	44 (3)	11 (5)
No	1,309 (97)	224 (95)
<b>Change of relationship quality between mother and father (1st wave)</b>		
Same	584 (43)	112 (48)
Worse	60 (4)	15 (6)
Better	709 (52)	108 (46)
<b>Relationship quality between mother and father (2nd wave)</b>		
Poor	0 (0)	0 (0)
Fair	0 (0)	0 (0)
Good	584 (43)	112 (48)
Very Good	60 (4)	15 (6)
Excellent	709 (52)	108 (46)
<b>Violence against mother, e.g., physical and verbal abuse (1st wave)</b>		
Yes	335 (25)	72 (31)
No	1,018 (75)	163 (69)

Table S3: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 3 (*continued*)

	Stable at W3 n (%)	Unstable at W3
<b>Violence against mother, e.g., physical and verbal abuse (2nd wave)</b>		
Yes	455 (34)	72 (31)
No	898 (66)	163 (69)
<b>Father was in jail (1st wave)</b>		
Yes	9 (1)	2 (1)
No	1,344 (99)	233 (99)
<b>Father was in jail (2nd wave)</b>		
Yes	6 (0)	4 (2)
No	1,347 (100)	231 (98)
<b>Public assistance, e.g, TANF or food stamps (1st wave)</b>		
Yes	275 (20)	81 (34)
No	1,078 (80)	154 (66)
<b>Public assistance, e.g, TANF or food stamps (2nd wave)</b>		
Yes	249 (18)	83 (35)
No	1,104 (82)	152 (65)
<b>Relatives provided financial assistance (1st wave)</b>		
Yes	394 (29)	102 (43)
No	959 (71)	133 (57)
<b>Relatives provided financial assistance (2nd wave)</b>		
Yes	401 (30)	87 (37)
No	952 (70)	148 (63)
<b>Poverty Categories based on Household income (1st wave)</b>		
More than 300%	544 (40)	51 (22)
Btw. 200-299%	237 (18)	39 (17)
Btw. 100-199%	295 (22)	63 (27)
Btw. 50-99%	150 (11)	34 (14)
Btw. 0-49%	127 (9)	48 (20)
<b>Poverty Categories based on Household income (2nd wave)</b>		
More than 300%	420 (31)	40 (17)
Btw. 200-299%	250 (18)	48 (20)
Btw. 100-199%	336 (25)	60 (26)
Btw. 50-99%	186 (14)	39 (17)
Btw. 0-49%	161 (12)	48 (20)
<b>Equivalized household income (1st wave)</b>		
min	0	0
max	94575.53	89974.39
mean (sd)	27,310.96 (24,424.96)	17,542.13 (17,911.69)
median	19006.58	12374.37
<b>Equivalized household income (2nd wave)</b>		
min	0	0
max	288675.1	139718.8
mean (sd)	23,471.78 (23,696.50)	16,242.67 (15,002.24)
median	17441.33	12500.00
<b>Housing wealth (2nd wave)</b>		
min	-200002	-15000
max	2000002	389000
mean (sd)	32,312.43 (98,235.09)	8,245.77 (34,785.99)
median	0	0
<b>Father's occupation (1st wave)</b>		
White collar, high skill	268 (20)	24 (10)

Table S3: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 3 (*continued*)

	Stable at W3 n (%)	Unstable at W3
Services, high skill	369 (27)	75 (32)
Manual blue collar	475 (35)	76 (32)
Other low skill	122 (9)	30 (13)
Self-employed	13 (1)	1 (0)
Unemployed	66 (5)	21 (9)
OLF	40 (3)	8 (3)
<b>Father's occupation (2nd wave)</b>		
White collar, high skill	338 (25)	26 (11)
Services, high skill	361 (27)	74 (31)
Manual blue collar	419 (31)	74 (31)
Other low skill	88 (7)	29 (12)
Self-employed	0 (0)	0 (0)
Unemployed	95 (7)	19 (8)
OLF	52 (4)	13 (6)
<b>Mother's occupation (2nd wave)</b>		
White collar, high skill	253 (19)	25 (11)
Services, high skill	437 (32)	108 (46)
Manual blue collar	32 (2)	4 (2)
Other low skill	23 (2)	3 (1)
Self-employed	0 (0)	0 (0)
Unemployed	176 (13)	54 (23)
OLF	432 (32)	41 (17)
<b>Did you change residence since previous wave? (1st wave)</b>		
Yes	553 (41)	116 (49)
No	800 (59)	119 (51)
<b>Did you change residence since previous wave? (2nd wave)</b>		
Yes	511 (38)	118 (50)
No	842 (62)	117 (50)
<b>Current living situation (1st wave)</b>		
Rent	776 (57)	171 (73)
Owned house/apt.	577 (43)	64 (27)
<b>Current living situation (2nd wave)</b>		
Rent	839 (62)	201 (86)
Owned house/apt.	514 (38)	34 (14)
<b>Neighborhood safety (1st wave)</b>		
Very unsafe	21 (2)	7 (3)
Unsafe	123 (9)	30 (13)
Safe	736 (54)	142 (60)
Very Safe	473 (35)	56 (24)
<b>Neighborhood safety (2nd wave)</b>		
Very unsafe	22 (2)	7 (3)
Unsafe	120 (9)	25 (11)
Safe	726 (54)	145 (62)
Very Safe	485 (36)	58 (25)

Table S4: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 4

	Stable at W4 n (%)	Unstable at W4
<b>Relatives/non-relatives in household (1st wave)</b>		
Yes	269 (24)	53 (32)
No	874 (76)	111 (68)
<b>Relatives/non-relatives in household (2nd wave)</b>		
Yes	206 (18)	40 (24)
No	937 (82)	124 (76)
<b>Relatives/non-relatives in household (3rd wave)</b>		
Yes	171 (15)	29 (18)
No	972 (85)	135 (82)
<b>Child lives with siblings from mother side (1st wave)</b>		
Yes	689 (60)	101 (62)
No	454 (40)	63 (38)
<b>Child lives with siblings from mother side (2nd wave)</b>		
Yes	1,050 (92)	149 (91)
No	93 (8)	15 (9)
<b>Child lives with siblings from mother side (3rd wave)</b>		
Yes	1,016 (89)	143 (87)
No	127 (11)	21 (13)
<b>Multipartner Fertility (2nd wave)</b>		
Yes	269 (24)	62 (38)
No	874 (76)	102 (62)
<b>Multipartner Fertility (3rd wave)</b>		
Yes	277 (24)	64 (39)
No	866 (76)	100 (61)
<b>Mother's general health (1st wave)</b>		
Fair or poor	66 (6)	15 (9)
Good	1,077 (94)	149 (91)
<b>Mother's general health (2nd wave)</b>		
Fair or poor	104 (9)	20 (12)
Good	1,039 (91)	144 (88)
<b>Mother's general health (3rd wave)</b>		
Fair or poor	100 (9)	20 (12)
Good	1,043 (91)	144 (88)
<b>Mother meets depression criteria (2nd wave)</b>		
Yes	114 (10)	27 (16)
No	1,029 (90)	137 (84)
<b>Mother meets depression criteria (3rd wave)</b>		
Yes	167 (15)	41 (25)
No	976 (85)	123 (75)
<b>Mother had alcohol problems (1st wave)</b>		
Yes	15 (1)	3 (2)
No	1,128 (99)	161 (98)
<b>Mother had alcohol problems (2nd wave)</b>		
Yes	7 (1)	1 (1)
No	1,136 (99)	163 (99)
<b>Mother had alcohol problems (3rd wave)</b>		
Yes	144 (13)	19 (12)
No	999 (87)	145 (88)
<b>Father had alcohol problems (1st wave)</b>		

Table S4: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 4 (*continued*)

	Stable at W4 n (%)	Unstable at W4
Yes	39 (3)	8 (5)
No	1,104 (97)	156 (95)
<b>Father had alcohol problems (2nd wave)</b>		
Yes	36 (3)	5 (3)
No	1,107 (97)	159 (97)
<b>Father had alcohol problems (3rd wave)</b>		
Yes	260 (23)	33 (20)
No	883 (77)	131 (80)
<b>Change of relationship quality between mother and father (1st wave)</b>		
Same	503 (44)	69 (42)
Worse	43 (4)	14 (9)
<b>Relationship quality between mother and father (2nd wave)</b>		
Better	597 (52)	81 (49)
Poor	5 (0)	3 (2)
Fair	41 (4)	9 (5)
Good	150 (13)	44 (27)
Very Good	403 (35)	53 (32)
Excellent	544 (48)	55 (34)
<b>Relationship quality between mother and father (3rd wave)</b>		
Poor	2 (0)	4 (2)
Fair	37 (3)	16 (10)
Good	180 (16)	46 (28)
Very Good	460 (40)	64 (39)
Excellent	464 (41)	34 (21)
<b>Violence against mother, e.g., physical and verbal abuse (1st wave)</b>		
Yes	277 (24)	51 (31)
No	866 (76)	113 (69)
<b>Violence against mother, e.g., physical and verbal abuse (2nd wave)</b>		
Yes	374 (33)	68 (41)
No	769 (67)	96 (59)
<b>Violence against mother, e.g., physical and verbal abuse (3rd wave)</b>		
Yes	368 (32)	69 (42)
No	775 (68)	95 (58)
<b>Father was in jail (1st wave)</b>		
Yes	7 (1)	1 (1)
No	1,136 (99)	163 (99)
<b>Father was in jail (2nd wave)</b>		
Yes	5 (0)	1 (1)
No	1,138 (100)	163 (99)
<b>Father was in jail (3rd wave)</b>		
Yes	8 (1)	5 (3)
No	1,135 (99)	159 (97)
<b>Public assistance, e.g, TANF or food stamps (1st wave)</b>		
Yes	209 (18)	49 (30)
No	934 (82)	115 (70)
<b>Public assistance, e.g, TANF or food stamps (2nd wave)</b>		
Yes	183 (16)	56 (34)
No	960 (84)	108 (66)
<b>Public assistance, e.g, TANF or food stamps (3rd wave)</b>		
Yes	200 (17)	52 (32)

Table S4: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 4 (*continued*)

	Stable at W4 n (%)	Unstable at W4
No	943 (83)	112 (68)
<b>Relatives provided financial assistance (1st wave)</b>		
Yes	304 (27)	67 (41)
No	839 (73)	97 (59)
<b>Relatives provided financial assistance (2nd wave)</b>		
Yes	314 (27)	66 (40)
No	829 (73)	98 (60)
<b>Relatives provided financial assistance (3rd wave)</b>		
Yes	220 (19)	49 (30)
No	923 (81)	115 (70)
<b>Poverty Categories based on Household income (1st wave)</b>		
More than 300%	492 (43)	39 (24)
Btw. 200-299%	191 (17)	36 (22)
Btw. 100-199%	251 (22)	35 (21)
Btw. 50-99%	117 (10)	30 (18)
Btw. 0-49%	92 (8)	24 (15)
<b>Poverty Categories based on Household income (2nd wave)</b>		
More than 300%	385 (34)	24 (15)
Btw. 200-299%	208 (18)	35 (21)
Btw. 100-199%	274 (24)	48 (29)
Btw. 50-99%	148 (13)	30 (18)
Btw. 0-49%	128 (11)	27 (16)
<b>Poverty Categories based on Household income (3rd wave)</b>		
More than 300%	430 (38)	42 (26)
Btw. 200-299%	204 (18)	30 (18)
Btw. 100-199%	248 (22)	44 (27)
Btw. 50-99%	150 (13)	22 (13)
Btw. 0-49%	111 (10)	26 (16)
<b>Equivalentized household income (1st wave)</b>		
min	0	0
max	94575.53	89974.39
mean (sd)	28,849.71 (25,086.07)	17,991.71 (16,346.87)
median	21213.2	15000.0
<b>Equivalentized household income (2nd wave)</b>		
min	0	0
max	288675.1	103923.0
mean (sd)	24,751.29 (24,769.03)	15,854.25 (14,020.82)
median	18158.5	11686.0
<b>Equivalentized household income (3rd wave)</b>		
min	0	0
max	577349.7	121243.6
mean (sd)	28,638.09 (33,155.21)	19,347.09 (15,711.86)
median	21000.00	16049.84
<b>Housing wealth (2nd wave)</b>		
min	-200002	-70002
max	2000002	245002
mean (sd)	36,315.76 (105,386.62)	9,102.46 (28,822.08)
median	0	0
<b>Housing wealth (3rd wave)</b>		
min	-38002	-15000
max	500002	200002

Table S4: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 4 (*continued*)

	Stable at W4 n (%)	Unstable at W4
mean (sd)	40,188.42 (89,074.98)	10,822.05 (27,507.08)
median	0	0
<b>Father's occupation (1st wave)</b>		
White collar, high skill	250 (22)	14 (9)
Services, high skill	309 (27)	43 (26)
Manual blue collar	398 (35)	61 (37)
Other low skill	95 (8)	24 (15)
Self-employed	9 (1)	4 (2)
Unemployed	50 (4)	11 (7)
OLF	32 (3)	7 (4)
<b>Father's occupation (2nd wave)</b>		
White collar, high skill	306 (27)	26 (16)
Services, high skill	304 (27)	45 (27)
Manual blue collar	346 (30)	55 (34)
Other low skill	74 (6)	10 (6)
Self-employed	0 (0)	0 (0)
Unemployed	71 (6)	20 (12)
OLF	42 (4)	8 (5)
<b>Father's occupation (3rd wave)</b>		
White collar, high skill	351 (31)	28 (17)
Services, high skill	253 (22)	32 (20)
Manual blue collar	364 (32)	57 (35)
Other low skill	78 (7)	21 (13)
Self-employed	0 (0)	0 (0)
Unemployed	60 (5)	13 (8)
OLF	37 (3)	13 (8)
<b>Mother's occupation (2nd wave)</b>		
White collar, high skill	229 (20)	18 (11)
Services, high skill	364 (32)	61 (37)
Manual blue collar	23 (2)	8 (5)
Other low skill	21 (2)	1 (1)
Self-employed	0 (0)	0 (0)
Unemployed	127 (11)	38 (23)
OLF	379 (33)	38 (23)
<b>Mother's occupation (3rd wave)</b>		
White collar, high skill	248 (22)	30 (18)
Services, high skill	379 (33)	61 (37)
Manual blue collar	20 (2)	4 (2)
Other low skill	8 (1)	1 (1)
Self-employed	0 (0)	0 (0)
Unemployed	136 (12)	31 (19)
OLF	352 (31)	37 (23)
<b>Did you change residence since previous wave? (1st wave)</b>		
Yes	456 (40)	78 (48)
No	687 (60)	86 (52)
<b>Did you change residence since previous wave? (2nd wave)</b>		
Yes	410 (36)	77 (47)
No	733 (64)	87 (53)
<b>Did you change residence since previous wave? (3rd wave)</b>		
Yes	474 (41)	79 (48)
No	669 (59)	85 (52)

Table S4: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 4 (*continued*)

	Stable at W4 n (%)	Unstable at W4
<b>Current living situation (1st wave)</b>		
Rent	632 (55)	113 (69)
Owned house/apt.	511 (45)	51 (31)
<b>Current living situation (2nd wave)</b>		
Rent	688 (60)	118 (72)
Owned house/apt.	455 (40)	46 (28)
<b>Current living situation (3rd wave)</b>		
Rent	606 (53)	117 (71)
Owned house/apt.	537 (47)	47 (29)
<b>Neighborhood safety (1st wave)</b>		
Very unsafe	16 (1)	4 (2)
Unsafe	96 (8)	16 (10)
Safe	620 (54)	96 (59)
Very Safe	411 (36)	48 (29)
<b>Neighborhood safety (2nd wave)</b>		
Very unsafe	16 (1)	5 (3)
Unsafe	96 (8)	16 (10)
Safe	603 (53)	98 (60)
Very Safe	428 (37)	45 (27)
<b>Neighborhood safety (3rd wave)</b>		
Very unsafe	32 (3)	6 (4)
Unsafe	46 (4)	10 (6)
Safe	79 (7)	17 (10)
Very Safe	986 (86)	131 (80)

Table S5: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 5

	Stable at W5 n (%)	Unstable at W5
<b>Relatives/non-relatives in household (1st wave)</b>		
Yes	207 (22)	47 (31)
No	725 (78)	106 (69)
<b>Relatives/non-relatives in household (2nd wave)</b>		
Yes	163 (17)	32 (21)
No	769 (83)	121 (79)
<b>Relatives/non-relatives in household (3rd wave)</b>		
Yes	126 (14)	35 (23)
No	806 (86)	118 (77)
<b>Relatives/non-relatives in household (4th wave)</b>		
Yes	120 (13)	24 (16)
No	812 (87)	129 (84)
<b>Child lives with siblings from mother side (1st wave)</b>		
Yes	554 (59)	100 (65)
No	378 (41)	53 (35)
<b>Child lives with siblings from mother side (2nd wave)</b>		
Yes	867 (93)	132 (86)
No	65 (7)	21 (14)
<b>Child lives with siblings from mother side (3rd wave)</b>		
Yes	841 (90)	126 (82)
No	91 (10)	27 (18)
<b>Child lives with siblings from mother side (4th wave)</b>		
Yes	810 (87)	117 (76)
No	122 (13)	36 (24)
<b>Multipartner Fertility (2nd wave)</b>		
Yes	196 (21)	57 (37)
No	736 (79)	96 (63)
<b>Multipartner Fertility (3rd wave)</b>		
Yes	202 (22)	58 (38)
No	730 (78)	95 (62)
<b>Multipartner Fertility (4th wave)</b>		
Yes	216 (23)	59 (39)
No	716 (77)	94 (61)
<b>Mother's general health (1st wave)</b>		
Fair or poor	46 (5)	15 (10)
Good	886 (95)	138 (90)
<b>Mother's general health (2nd wave)</b>		
Fair or poor	83 (9)	16 (10)
Good	849 (91)	137 (90)
<b>Mother's general health (3rd wave)</b>		
Fair or poor	73 (8)	18 (12)
Good	859 (92)	135 (88)
<b>Mother's general health (4th wave)</b>		
Fair or poor	70 (8)	18 (12)
Good	862 (92)	135 (88)
<b>Mother meets depression criteria (2nd wave)</b>		
Yes	85 (9)	21 (14)
No	847 (91)	132 (86)
<b>Mother meets depression criteria (3rd wave)</b>		

Table S5: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 5 (*continued*)

	Stable at W5 n (%)	Unstable at W5
Yes	123 (13)	33 (22)
No	809 (87)	120 (78)
<b>Mother meets depression criteria (4th wave)</b>		
Yes	103 (11)	26 (17)
No	829 (89)	127 (83)
<b>Mother had alcohol problems (1st wave)</b>		
Yes	9 (1)	3 (2)
No	923 (99)	150 (98)
<b>Mother had alcohol problems (2nd wave)</b>		
Yes	5 (1)	2 (1)
No	927 (99)	151 (99)
<b>Mother had alcohol problems (3rd wave)</b>		
Yes	110 (12)	24 (16)
No	822 (88)	129 (84)
<b>Mother had alcohol problems (4th wave)</b>		
Yes	20 (2)	5 (3)
No	912 (98)	148 (97)
<b>Father had alcohol problems (1st wave)</b>		
Yes	30 (3)	5 (3)
No	902 (97)	148 (97)
<b>Father had alcohol problems (2nd wave)</b>		
Yes	25 (3)	8 (5)
No	907 (97)	145 (95)
<b>Father had alcohol problems (3rd wave)</b>		
Yes	206 (22)	40 (26)
No	726 (78)	113 (74)
<b>Father had alcohol problems (4th wave)</b>		
Yes	11 (1)	4 (3)
No	921 (99)	149 (97)
<b>Change of relationship quality between mother and father (1st wave)</b>		
Same	408 (44)	64 (42)
Worse	32 (3)	9 (6)
Better	492 (53)	80 (52)
<b>Relationship quality between mother and father (2nd wave)</b>		
Poor	5 (1)	0 (0)
Fair	28 (3)	9 (6)
Good	110 (12)	31 (20)
Very Good	328 (35)	55 (36)
Excellent	461 (49)	58 (38)
<b>Relationship quality between mother and father (3rd wave)</b>		
Poor	0 (0)	1 (1)
Fair	22 (2)	12 (8)
Good	125 (13)	45 (29)
Very Good	378 (41)	58 (38)
Excellent	407 (44)	37 (24)
<b>Relationship quality between mother and father (4th wave)</b>		
Poor	6 (1)	7 (5)
Fair	23 (2)	21 (14)
Good	139 (15)	36 (24)
Very Good	353 (38)	50 (33)

Table S5: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 5 (*continued*)

	Stable at W5 n (%)	Unstable at W5
Excellent	411 (44)	39 (25)
<b>Violence against mother, e.g., physical and verbal abuse (1st wave)</b>		
Yes	213 (23)	51 (33)
No	719 (77)	102 (67)
<b>Violence against mother, e.g., physical and verbal abuse (2nd wave)</b>		
Yes	304 (33)	49 (32)
No	628 (67)	104 (68)
<b>Violence against mother, e.g., physical and verbal abuse (3rd wave)</b>		
Yes	295 (32)	56 (37)
No	637 (68)	97 (63)
<b>Violence against mother, e.g., physical and verbal abuse (4th wave)</b>		
Yes	294 (32)	58 (38)
No	638 (68)	95 (62)
<b>Father was in jail (1st wave)</b>		
Yes	4 (0)	2 (1)
No	928 (100)	151 (99)
<b>Father was in jail (2nd wave)</b>		
Yes	1 (0)	3 (2)
No	931 (100)	150 (98)
<b>Father was in jail (3rd wave)</b>		
Yes	4 (0)	4 (3)
No	928 (100)	149 (97)
<b>Father was in jail (4th wave)</b>		
Yes	4 (0)	8 (5)
No	928 (100)	145 (95)
<b>Public assistance, e.g, TANF or food stamps (1st wave)</b>		
Yes	144 (15)	49 (32)
No	788 (85)	104 (68)
<b>Public assistance, e.g, TANF or food stamps (2nd wave)</b>		
Yes	129 (14)	38 (25)
No	803 (86)	115 (75)
<b>Public assistance, e.g, TANF or food stamps (3rd wave)</b>		
Yes	147 (16)	38 (25)
No	785 (84)	115 (75)
<b>Public assistance, e.g, TANF or food stamps (4th wave)</b>		
Yes	132 (14)	40 (26)
No	800 (86)	113 (74)
<b>Relatives provided financial assistance (1st wave)</b>		
Yes	233 (25)	53 (35)
No	699 (75)	100 (65)
<b>Relatives provided financial assistance (2nd wave)</b>		
Yes	246 (26)	50 (33)
No	686 (74)	103 (67)
<b>Relatives provided financial assistance (3rd wave)</b>		
Yes	164 (18)	37 (24)
No	768 (82)	116 (76)
<b>Relatives provided financial assistance (4th wave)</b>		
Yes	184 (20)	42 (27)
No	748 (80)	111 (73)

Table S5: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 5 (*continued*)

	Stable at W5 n (%)	Unstable at W5
<b>Poverty Categories based on Household income (1st wave)</b>		
More than 300%	433 (46)	44 (29)
Btw. 200-299%	165 (18)	18 (12)
Btw. 100-199%	180 (19)	47 (31)
Btw. 50-99%	92 (10)	21 (14)
Btw. 0-49%	62 (7)	23 (15)
<b>Poverty Categories based on Household income (2nd wave)</b>		
More than 300%	343 (37)	30 (20)
Btw. 200-299%	170 (18)	28 (18)
Btw. 100-199%	204 (22)	48 (31)
Btw. 50-99%	118 (13)	21 (14)
Btw. 0-49%	97 (10)	26 (17)
<b>Poverty Categories based on Household income (3rd wave)</b>		
More than 300%	381 (41)	34 (22)
Btw. 200-299%	162 (17)	32 (21)
Btw. 100-199%	191 (20)	36 (24)
Btw. 50-99%	117 (13)	28 (18)
Btw. 0-49%	81 (9)	23 (15)
<b>Poverty Categories based on Household income (4th wave)</b>		
More than 300%	367 (39)	43 (28)
Btw. 200-299%	171 (18)	30 (20)
Btw. 100-199%	206 (22)	42 (27)
Btw. 50-99%	102 (11)	19 (12)
Btw. 0-49%	86 (9)	19 (12)
<b>Equivalized household income (1st wave)</b>		
min	0	0
max	94575.53	89974.39
mean (sd)	31,037.28 (25,904.28)	19,558.24 (19,016.54)
median	21250	12500
<b>Equivalized household income (2nd wave)</b>		
min	0.0000	662.3233
max	288675.1	178885.4
mean (sd)	26,091.06 (25,548.64)	18,875.92 (20,739.49)
median	19838.70	13097.19
<b>Equivalized household income (3rd wave)</b>		
min	0.0000	115.4701
max	577349.7	100000.0
mean (sd)	30,523.89 (35,245.85)	18,981.81 (16,093.22)
median	22500.00	16212.83
<b>Equivalized household income (4th wave)</b>		
min	0	0
max	357770.88	86602.54
mean (sd)	31,009.22 (31,888.42)	21,839.15 (16,377.84)
median	23094.73	17888.54
<b>Housing wealth (2nd wave)</b>		
min	-200002	-19000
max	2000002	750000
mean (sd)	39,044.95 (110,332.70)	26,353.03 (84,658.07)
median	0	0
<b>Housing wealth (3rd wave)</b>		
min	-38002	-9000

Table S5: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 5 (*continued*)

	Stable at W5 n (%)	Unstable at W5
max	500002	500002
mean (sd)	42,235.46 (89,504.43)	32,822.95 (90,532.56)
median	0	0
<b>Housing wealth (4th wave)</b>		
min	-13000	-3000
max	400002	400002
mean (sd)	37,344.19 (82,750.94)	23,843.16 (73,488.42)
median	0	0
<b>Father's occupation (1st wave)</b>		
White collar, high skill	226 (24)	17 (11)
Services, high skill	250 (27)	45 (29)
Manual blue collar	321 (34)	55 (36)
Other low skill	76 (8)	15 (10)
Self-employed	8 (1)	0 (0)
Unemployed	32 (3)	11 (7)
OLF	19 (2)	10 (7)
<b>Father's occupation (2nd wave)</b>		
White collar, high skill	274 (29)	24 (16)
Services, high skill	240 (26)	51 (33)
Manual blue collar	277 (30)	47 (31)
Other low skill	61 (7)	8 (5)
Self-employed	0 (0)	0 (0)
Unemployed	50 (5)	14 (9)
OLF	30 (3)	9 (6)
<b>Father's occupation (3rd wave)</b>		
White collar, high skill	308 (33)	30 (20)
Services, high skill	214 (23)	30 (20)
Manual blue collar	283 (30)	55 (36)
Other low skill	61 (7)	12 (8)
Self-employed	0 (0)	0 (0)
Unemployed	43 (5)	13 (8)
OLF	23 (2)	13 (8)
<b>Father's occupation (4th wave)</b>		
White collar, high skill	322 (35)	29 (19)
Services, high skill	190 (20)	30 (20)
Manual blue collar	284 (30)	57 (37)
Other low skill	66 (7)	16 (10)
Self-employed	0 (0)	0 (0)
Unemployed	41 (4)	7 (5)
OLF	29 (3)	14 (9)
<b>Mother's occupation (2nd wave)</b>		
White collar, high skill	196 (21)	24 (16)
Services, high skill	297 (32)	46 (30)
Manual blue collar	11 (1)	9 (6)
Other low skill	17 (2)	3 (2)
Self-employed	0 (0)	0 (0)
Unemployed	93 (10)	27 (18)
OLF	318 (34)	44 (29)
<b>Mother's occupation (3rd wave)</b>		
White collar, high skill	219 (23)	22 (14)
Services, high skill	297 (32)	62 (41)
Manual blue collar	14 (2)	6 (4)

Table S5: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 5 (*continued*)

	Stable at W5 n (%)	Unstable at W5
Other low skill	6 (1)	0 (0)
Self-employed	0 (0)	0 (0)
Unemployed	101 (11)	25 (16)
OLF	295 (32)	38 (25)
<b>Mother's occupation (4th wave)</b>		
White collar, high skill	216 (23)	28 (18)
Services, high skill	340 (36)	69 (45)
Manual blue collar	18 (2)	2 (1)
Other low skill	7 (1)	2 (1)
Self-employed	0 (0)	0 (0)
Unemployed	83 (9)	29 (19)
OLF	268 (29)	23 (15)
<b>Did you change residence since previous wave? (1st wave)</b>		
Yes	358 (38)	69 (45)
No	574 (62)	84 (55)
<b>Did you change residence since previous wave? (2nd wave)</b>		
Yes	322 (35)	64 (42)
No	610 (65)	89 (58)
<b>Did you change residence since previous wave? (3rd wave)</b>		
Yes	372 (40)	75 (49)
No	560 (60)	78 (51)
<b>Did you change residence since previous wave? (4th wave)</b>		
Yes	354 (38)	57 (37)
No	578 (62)	96 (63)
<b>Current living situation (1st wave)</b>		
Rent	501 (54)	93 (61)
Owned house/apt.	431 (46)	60 (39)
<b>Current living situation (2nd wave)</b>		
Rent	546 (59)	98 (64)
Owned house/apt.	386 (41)	55 (36)
<b>Current living situation (3rd wave)</b>		
Rent	472 (51)	95 (62)
Owned house/apt.	460 (49)	58 (38)
<b>Current living situation (4th wave)</b>		
Rent	539 (58)	113 (74)
Owned house/apt.	393 (42)	40 (26)
<b>Neighborhood safety (1st wave)</b>		
Very unsafe	10 (1)	5 (3)
Unsafe	64 (7)	24 (16)
Safe	506 (54)	82 (54)
Very Safe	352 (38)	42 (27)
<b>Neighborhood safety (2nd wave)</b>		
Very unsafe	12 (1)	3 (2)
Unsafe	72 (8)	19 (12)
Safe	498 (53)	77 (50)
Very Safe	350 (38)	54 (35)
<b>Neighborhood safety (3rd wave)</b>		
Very unsafe	25 (3)	6 (4)
Unsafe	36 (4)	6 (4)
Safe	70 (8)	7 (5)
Very Safe	801 (86)	134 (88)

Table S5: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 5 (*continued*)

	Stable at W5 n (%)	Unstable at W5
<b>Neighborhood safety (4th wave)</b>		
Yes	85 (9)	11 (7)
No	847 (91)	142 (93)

Table S6: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 6

	Stable at W6 n (%)	Unstable at W6
<b>Relatives/non-relatives in household (1st wave)</b>		
Yes	145 (20)	45 (26)
No	570 (80)	125 (74)
<b>Relatives/non-relatives in household (2nd wave)</b>		
Yes	113 (16)	39 (23)
No	602 (84)	131 (77)
<b>Relatives/non-relatives in household (3rd wave)</b>		
Yes	84 (12)	29 (17)
No	631 (88)	141 (83)
<b>Relatives/non-relatives in household (4th wave)</b>		
Yes	88 (12)	20 (12)
No	627 (88)	150 (88)
<b>Relatives/non-relatives in household (5th wave)</b>		
Yes	68 (10)	19 (11)
No	647 (90)	151 (89)
<b>Child lives with siblings from mother side (1st wave)</b>		
Yes	418 (58)	104 (61)
No	297 (42)	66 (39)
<b>Child lives with siblings from mother side (2nd wave)</b>		
Yes	660 (92)	162 (95)
No	55 (8)	8 (5)
<b>Child lives with siblings from mother side (3rd wave)</b>		
Yes	646 (90)	149 (88)
No	69 (10)	21 (12)
<b>Child lives with siblings from mother side (4th wave)</b>		
Yes	629 (88)	144 (85)
No	86 (12)	26 (15)
<b>Child lives with siblings from mother side (5th wave)</b>		
Yes	563 (79)	120 (71)
No	152 (21)	50 (29)
<b>Multipartner Fertility (2nd wave)</b>		
Yes	144 (20)	38 (22)
No	571 (80)	132 (78)
<b>Multipartner Fertility (3rd wave)</b>		
Yes	146 (20)	41 (24)
No	569 (80)	129 (76)
<b>Multipartner Fertility (4th wave)</b>		
Yes	158 (22)	42 (25)
No	557 (78)	128 (75)
<b>Multipartner Fertility (5th wave)</b>		
Yes	163 (23)	45 (26)
No	552 (77)	125 (74)
<b>Mother's general health (1st wave)</b>		
Fair or poor	31 (4)	13 (8)
Good	684 (96)	157 (92)
<b>Mother's general health (2nd wave)</b>		
Fair or poor	61 (9)	17 (10)
Good	654 (91)	153 (90)
<b>Mother's general health (3rd wave)</b>		

Table S6: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 6 (*continued*)

	Stable at W6 n (%)	Unstable at W6
Fair or poor	48 (7)	17 (10)
Good	667 (93)	153 (90)
<b>Mother's general health (4th wave)</b>		
Fair or poor	57 (8)	9 (5)
Good	658 (92)	161 (95)
<b>Mother's general health (5th wave)</b>		
Fair or poor	69 (10)	21 (12)
Good	646 (90)	149 (88)
<b>Mother meets depression criteria (2nd wave)</b>		
Yes	66 (9)	14 (8)
No	649 (91)	156 (92)
<b>Mother meets depression criteria (3rd wave)</b>		
Yes	91 (13)	25 (15)
No	624 (87)	145 (85)
<b>Mother meets depression criteria (4th wave)</b>		
Yes	76 (11)	21 (12)
No	639 (89)	149 (88)
<b>Mother meets depression criteria (5th wave)</b>		
Yes	89 (12)	16 (9)
No	626 (88)	154 (91)
<b>Mother had alcohol problems (1st wave)</b>		
Yes	5 (1)	3 (2)
No	710 (99)	167 (98)
<b>Mother had alcohol problems (2nd wave)</b>		
Yes	2 (0)	3 (2)
No	713 (100)	167 (98)
<b>Mother had alcohol problems (3rd wave)</b>		
Yes	82 (11)	16 (9)
No	633 (89)	154 (91)
<b>Mother had alcohol problems (4th wave)</b>		
Yes	15 (2)	3 (2)
No	700 (98)	167 (98)
<b>Mother had alcohol problems (5th wave)</b>		
Yes	9 (1)	1 (1)
No	706 (99)	169 (99)
<b>Father had alcohol problems (1st wave)</b>		
Yes	22 (3)	3 (2)
No	693 (97)	167 (98)
<b>Father had alcohol problems (2nd wave)</b>		
Yes	17 (2)	7 (4)
No	698 (98)	163 (96)
<b>Father had alcohol problems (3rd wave)</b>		
Yes	156 (22)	39 (23)
No	559 (78)	131 (77)
<b>Father had alcohol problems (4th wave)</b>		
Yes	7 (1)	4 (2)
No	708 (99)	166 (98)
<b>Father had alcohol problems (5th wave)</b>		
Yes	11 (2)	7 (4)
No	704 (98)	163 (96)

Table S6: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 6 (*continued*)

	Stable at W6 n (%)	Unstable at W6
<b>Change of relationship quality between mother and father (1st wave)</b>		
Same	313 (44)	74 (44)
Worse	24 (3)	7 (4)
Better	378 (53)	89 (52)
<b>Relationship quality between mother and father (2nd wave)</b>		
Poor	4 (1)	1 (1)
Fair	18 (3)	9 (5)
Good	79 (11)	23 (14)
Very Good	259 (36)	58 (34)
Excellent	355 (50)	79 (46)
<b>Relationship quality between mother and father (3rd wave)</b>		
Poor	0 (0)	0 (0)
Fair	15 (2)	4 (2)
Good	92 (13)	28 (16)
Very Good	285 (40)	72 (42)
Excellent	323 (45)	66 (39)
<b>Relationship quality between mother and father (4th wave)</b>		
Poor	3 (0)	3 (2)
Fair	17 (2)	5 (3)
Good	94 (13)	32 (19)
Very Good	276 (39)	60 (35)
Excellent	325 (45)	70 (41)
<b>Relationship quality between mother and father (5th wave)</b>		
Poor	1 (0)	5 (3)
Fair	20 (3)	9 (5)
Good	118 (17)	43 (25)
Very Good	296 (41)	61 (36)
Excellent	280 (39)	52 (31)
<b>Violence against mother, e.g., physical and verbal abuse (1st wave)</b>		
Yes	167 (23)	43 (25)
No	548 (77)	127 (75)
<b>Violence against mother, e.g., physical and verbal abuse (2nd wave)</b>		
Yes	232 (32)	57 (34)
No	483 (68)	113 (66)
<b>Violence against mother, e.g., physical and verbal abuse (3rd wave)</b>		
Yes	224 (31)	59 (35)
No	491 (69)	111 (65)
<b>Violence against mother, e.g., physical and verbal abuse (4th wave)</b>		
Yes	225 (31)	61 (36)
No	490 (69)	109 (64)
<b>Violence against mother, e.g., physical and verbal abuse (5th wave)</b>		
Yes	187 (26)	52 (31)
No	528 (74)	118 (69)
<b>Father was in jail (1st wave)</b>		
Yes	3 (0)	1 (1)
No	712 (100)	169 (99)
<b>Father was in jail (2nd wave)</b>		
Yes	1 (0)	0 (0)
No	714 (100)	170 (100)
<b>Father was in jail (3rd wave)</b>		

Table S6: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 6 (*continued*)

	Stable at W6 n (%)	Unstable at W6
Yes	4 (1)	0 (0)
No	711 (99)	170 (100)
<b>Father was in jail (4th wave)</b>		
Yes	3 (0)	1 (1)
No	712 (100)	169 (99)
<b>Father was in jail (5th wave)</b>		
Yes	1 (0)	0 (0)
No	714 (100)	170 (100)
<b>Public assistance, e.g, TANF or food stamps (1st wave)</b>		
Yes	93 (13)	36 (21)
No	622 (87)	134 (79)
<b>Public assistance, e.g, TANF or food stamps (2nd wave)</b>		
Yes	85 (12)	32 (19)
No	630 (88)	138 (81)
<b>Public assistance, e.g, TANF or food stamps (3rd wave)</b>		
Yes	96 (13)	37 (22)
No	619 (87)	133 (78)
<b>Public assistance, e.g, TANF or food stamps (4th wave)</b>		
Yes	93 (13)	27 (16)
No	622 (87)	143 (84)
<b>Public assistance, e.g, TANF or food stamps (5th wave)</b>		
Yes	143 (20)	52 (31)
No	572 (80)	118 (69)
<b>Relatives provided financial assistance (1st wave)</b>		
Yes	167 (23)	45 (26)
No	548 (77)	125 (74)
<b>Relatives provided financial assistance (2nd wave)</b>		
Yes	193 (27)	42 (25)
No	522 (73)	128 (75)
<b>Relatives provided financial assistance (3rd wave)</b>		
Yes	129 (18)	31 (18)
No	586 (82)	139 (82)
<b>Relatives provided financial assistance (4th wave)</b>		
Yes	139 (19)	39 (23)
No	576 (81)	131 (77)
<b>Relatives provided financial assistance (5th wave)</b>		
Yes	157 (22)	41 (24)
No	558 (78)	129 (76)
<b>Poverty Categories based on Household income (1st wave)</b>		
More than 300%	364 (51)	61 (36)
Btw. 200-299%	114 (16)	43 (25)
Btw. 100-199%	136 (19)	32 (19)
Btw. 50-99%	65 (9)	20 (12)
Btw. 0-49%	36 (5)	14 (8)
<b>Poverty Categories based on Household income (2nd wave)</b>		
More than 300%	295 (41)	41 (24)
Btw. 200-299%	125 (17)	36 (21)
Btw. 100-199%	140 (20)	52 (31)
Btw. 50-99%	89 (12)	21 (12)
Btw. 0-49%	66 (9)	20 (12)

Table S6: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 6 (*continued*)

	Stable at W6 n (%)	Unstable at W6
<b>Poverty Categories based on Household income (3rd wave)</b>		
More than 300%	318 (44)	52 (31)
Btw. 200-299%	122 (17)	34 (20)
Btw. 100-199%	144 (20)	38 (22)
Btw. 50-99%	76 (11)	31 (18)
Btw. 0-49%	55 (8)	15 (9)
<b>Poverty Categories based on Household income (4th wave)</b>		
More than 300%	302 (42)	55 (32)
Btw. 200-299%	128 (18)	35 (21)
Btw. 100-199%	151 (21)	43 (25)
Btw. 50-99%	78 (11)	18 (11)
Btw. 0-49%	56 (8)	19 (11)
<b>Poverty Categories based on Household income (5th wave)</b>		
More than 300%	299 (42)	52 (31)
Btw. 200-299%	122 (17)	24 (14)
Btw. 100-199%	156 (22)	59 (35)
Btw. 50-99%	77 (11)	24 (14)
Btw. 0-49%	61 (9)	11 (6)
<b>Equivalized household income (1st wave)</b>		
min	0.000	1530.931
max	94575.53	94575.53
mean (sd)	32,957.43 (26,310.14)	27,123.76 (24,486.69)
median	24537.39	19006.58
<b>Equivalized household income (2nd wave)</b>		
min	1.732051	167.381799
max	288675.1	200000.0
mean (sd)	27,760.63 (26,388.27)	21,875.19 (23,094.84)
median	20581.30	15876.08
<b>Equivalized household income (3rd wave)</b>		
min	0.0000	604.6328
max	577349.7	175000.0
mean (sd)	32,854.49 (38,112.39)	24,002.83 (22,173.02)
median	24494.90	18592.25
<b>Equivalized household income (4th wave)</b>		
min	0	0
max	357770.9	200000.0
mean (sd)	33,005.43 (33,877.82)	26,338.48 (24,334.47)
median	24596.75	20576.42
<b>Equivalized household income (5th wave)</b>		
min	-1.341641	0.000000
max	450000	300000
mean (sd)	36,271.92 (36,168.65)	28,846.40 (29,768.30)
median	27280.03	20268.51
<b>Housing wealth (2nd wave)</b>		
min	-200002	-30002
max	2000002	600000
mean (sd)	44,106.27 (119,000.49)	22,011.24 (69,615.42)
median	0	0
<b>Housing wealth (3rd wave)</b>		
min	-9000	-38002
max	500002	500000

Table S6: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 6 (*continued*)

	Stable at W6 n (%)	Unstable at W6
mean (sd)	47,462.43 (94,566.75)	25,617.08 (63,626.02)
median	0	0
<b>Housing wealth (4th wave)</b>		
min	-13000	-13000
max	400002	400002
mean (sd)	40,446.08 (86,051.76)	28,046.13 (71,092.92)
median	0	0
<b>Housing wealth (5th wave)</b>		
min	-300002	-300002
max	3000002	1200002
mean (sd)	115,198.68 (234,243.98)	69,250.06 (152,705.71)
median	33000	0
<b>Father's occupation (1st wave)</b>		
White collar, high skill	192 (27)	28 (16)
Services, high skill	186 (26)	54 (32)
Manual blue collar	236 (33)	66 (39)
Other low skill	57 (8)	14 (8)
Self-employed	6 (1)	1 (1)
Unemployed	24 (3)	6 (4)
OLF	14 (2)	1 (1)
<b>Father's occupation (2nd wave)</b>		
White collar, high skill	230 (32)	33 (19)
Services, high skill	179 (25)	54 (32)
Manual blue collar	210 (29)	48 (28)
Other low skill	45 (6)	13 (8)
Self-employed	0 (0)	0 (0)
Unemployed	34 (5)	14 (8)
OLF	17 (2)	8 (5)
<b>Father's occupation (3rd wave)</b>		
White collar, high skill	245 (34)	57 (34)
Services, high skill	175 (24)	30 (18)
Manual blue collar	204 (29)	57 (34)
Other low skill	46 (6)	12 (7)
Self-employed	0 (0)	0 (0)
Unemployed	31 (4)	9 (5)
OLF	14 (2)	5 (3)
<b>Father's occupation (4th wave)</b>		
White collar, high skill	266 (37)	46 (27)
Services, high skill	143 (20)	41 (24)
Manual blue collar	207 (29)	59 (35)
Other low skill	51 (7)	9 (5)
Self-employed	0 (0)	0 (0)
Unemployed	27 (4)	8 (5)
OLF	21 (3)	7 (4)
<b>Father's occupation (5th wave)</b>		
White collar, high skill	258 (36)	40 (24)
Services, high skill	151 (21)	35 (21)
Manual blue collar	203 (28)	46 (27)
Other low skill	35 (5)	14 (8)
Self-employed	0 (0)	0 (0)
Unemployed	45 (6)	25 (15)
OLF	23 (3)	10 (6)

Table S6: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 6 (*continued*)

	Stable at W6 n (%)	Unstable at W6
<b>Mother's occupation (2nd wave)</b>		
White collar, high skill	159 (22)	32 (19)
Services, high skill	229 (32)	54 (32)
Manual blue collar	10 (1)	1 (1)
Other low skill	15 (2)	1 (1)
Self-employed	0 (0)	0 (0)
Unemployed	63 (9)	22 (13)
OLF	239 (33)	60 (35)
<b>Mother's occupation (3rd wave)</b>		
White collar, high skill	176 (25)	38 (22)
Services, high skill	225 (31)	58 (34)
Manual blue collar	10 (1)	3 (2)
Other low skill	5 (1)	1 (1)
Self-employed	0 (0)	0 (0)
Unemployed	63 (9)	24 (14)
OLF	236 (33)	46 (27)
<b>Mother's occupation (4th wave)</b>		
White collar, high skill	175 (24)	33 (19)
Services, high skill	251 (35)	74 (44)
Manual blue collar	13 (2)	4 (2)
Other low skill	6 (1)	0 (0)
Self-employed	0 (0)	0 (0)
Unemployed	56 (8)	17 (10)
OLF	214 (30)	42 (25)
<b>Mother's occupation (5th wave)</b>		
White collar, high skill	219 (31)	46 (27)
Services, high skill	233 (33)	64 (38)
Manual blue collar	18 (3)	3 (2)
Other low skill	7 (1)	1 (1)
Self-employed	0 (0)	0 (0)
Unemployed	65 (9)	23 (14)
OLF	173 (24)	33 (19)
<b>Did you change residence since previous wave? (1st wave)</b>		
Yes	255 (36)	75 (44)
No	460 (64)	95 (56)
<b>Did you change residence since previous wave? (2nd wave)</b>		
Yes	226 (32)	76 (45)
No	489 (68)	94 (55)
<b>Did you change residence since previous wave? (3rd wave)</b>		
Yes	277 (39)	71 (42)
No	438 (61)	99 (58)
<b>Did you change residence since previous wave? (4th wave)</b>		
Yes	261 (37)	77 (45)
No	454 (63)	93 (55)
<b>Did you change residence since previous wave? (5th wave)</b>		
Yes	277 (39)	78 (46)
No	438 (61)	92 (54)
<b>Current living situation (1st wave)</b>		
Rent	366 (51)	101 (59)
Owned house/apt.	349 (49)	69 (41)
<b>Current living situation (2nd wave)</b>		

Table S6: Descriptive Statistics for Time-varying Analytical Variables in the FFCWS at Wave 6 (*continued*)

	Stable at W6 n (%)	Unstable at W6
Rent	395 (55)	113 (66)
Owned house/apt.	320 (45)	57 (34)
<b>Current living situation (3rd wave)</b>		
Rent	338 (47)	101 (59)
Owned house/apt.	377 (53)	69 (41)
<b>Current living situation (4th wave)</b>		
Rent	405 (57)	101 (59)
Owned house/apt.	310 (43)	69 (41)
<b>Current living situation (5th wave)</b>		
Rent	313 (44)	92 (54)
Owned house/apt.	402 (56)	78 (46)
<b>Neighborhood safety (1st wave)</b>		
Very unsafe	6 (1)	3 (2)
Unsafe	48 (7)	13 (8)
Safe	383 (54)	95 (56)
Very Safe	278 (39)	59 (35)
<b>Neighborhood safety (2nd wave)</b>		
Very unsafe	6 (1)	4 (2)
Unsafe	51 (7)	18 (11)
Safe	381 (53)	90 (53)
Very Safe	277 (39)	58 (34)
<b>Neighborhood safety (3rd wave)</b>		
Very unsafe	17 (2)	8 (5)
Unsafe	27 (4)	7 (4)
Safe	52 (7)	16 (9)
Very Safe	619 (87)	139 (82)
<b>Neighborhood safety (4th wave)</b>		
Yes	59 (8)	21 (12)
No	656 (92)	149 (88)
<b>Neighborhood safety (5th wave)</b>		
Yes	80 (11)	27 (16)
No	635 (89)	143 (84)

**Supplementary Materials - Chapter 4**  
**Descriptive statistics**

Table S7: Descriptive Statistics by Parents' Socioeconomic Status, NEPS SC1

	All	Socioeconomic Status Latent Classes			
	n (%)	Very Low (N = 221)	Low (N = 553)	Medium (N = 261)	High (N = 857)
<b>Baby's gender assigned at birth</b>					
Boy	981 (52)	117 (53)	293 (53)	145 (56)	426 (50)
Girl	911 (48)	104 (47)	260 (47)	116 (44)	431 (50)
<b>Birth Order</b>					
First	631 (33)	57 (26)	161 (29)	101 (39)	312 (36)
Second or later	1,261 (67)	164 (74)	392 (71)	160 (61)	545 (64)
<b>Premature birth</b>					
Yes	112 (6)	16 (7)	29 (5)	13 (5)	54 (6)
No	1,780 (94)	205 (93)	524 (95)	248 (95)	803 (94)
<b>Low birthweight (&lt; 2500g)</b>					
Yes	108 (6)	20 (9)	30 (5)	12 (5)	46 (5)
No	1,784 (94)	201 (91)	523 (95)	249 (95)	811 (95)
<b>Smoke while pregnant</b>					
Yes, regularly, now and then	194 (10)	112 (51)	62 (11)	6 (2)	14 (2)
No, never	1,698 (90)	109 (49)	491 (89)	255 (98)	843 (98)
<b>Drank alcohol while pregnant</b>					
Yes, regularly, now and then	129 (7)	10 (5)	30 (5)	16 (6)	73 (9)
No, never	1,763 (93)	211 (95)	523 (95)	245 (94)	784 (91)
<b>Months breastfed</b>					
Not breastfed	212 (11)	75 (34)	76 (14)	15 (6)	46 (5)
Btw. 1-3 months	284 (15)	71 (32)	117 (21)	25 (10)	71 (8)
Btw. 4-6 months	1,135 (60)	61 (28)	302 (55)	165 (63)	607 (71)
More than 6 months	261 (14)	14 (6)	58 (10)	56 (21)	133 (16)
<b>Mother's feelings of depression</b>					
Never	670 (35)	49 (22)	182 (33)	87 (33)	352 (41)
Seldom	723 (38)	78 (35)	211 (38)	111 (43)	323 (38)
Sometimes	355 (19)	61 (28)	114 (21)	50 (19)	130 (15)
Often/always	144 (8)	33 (15)	46 (8)	13 (5)	52 (6)
<b>Mother's age</b>					
Mean (sd)	32.46 (5.11)	27.43 (5.81)	31.18 (4.99)	32.43 (4.24)	34.59 (3.92)
Median	32	27	31	32	34
<b>Family structure at birth</b>					
Two biological parent	1,717 (91)	122 (55)	519 (94)	231 (89)	845 (99)
Two parents (stepfather)	16 (1)	5 (2)	6 (1)	0 (0)	5 (1)
Lone mother	159 (8)	94 (43)	28 (5)	30 (11)	7 (1)

Table S7: Descriptive Statistics by Parents' Socioeconomic Status, NEPS SC1 (*continued*)

	All	Socioeconomic Status Latent Classes			
	n (%)	Very Low (N = 221)	Low (N = 553)	Medium (N = 261)	High (N = 857)
<b>Family owns house they live in</b>					
Yes	815 (43)	9 (4)	238 (43)	59 (23)	509 (59)
No	1,077 (57)	212 (96)	315 (57)	202 (77)	348 (41)
<b>Residence location in Germany</b>					
East	505 (27)	113 (51)	143 (26)	123 (47)	126 (15)
West	1,387 (73)	108 (49)	410 (74)	138 (53)	731 (85)
<b>Cultural capital index</b>					
Mean (sd)	8.69 (2.98)	7.21 (2.18)	8.02 (2.58)	9.64 (3.11)	9.22 (3.13)
Median	8	7	7	9	9
<b>Number of adults in household</b>					
One	140 (7)	83 (38)	22 (4)	28 (11)	7 (1)
Two	1,694 (90)	126 (57)	502 (91)	230 (88)	836 (98)
More than two	58 (3)	12 (5)	29 (5)	3 (1)	14 (2)
<b>ELFRA-2P wave 3 sum score</b>					
Mean (sd)	147.42 (63.80)	105.00 (66.20)	142.25 (66.59)	157.57 (58.01)	158.60 (57.85)
Median	158	99	153	167	166
<b>PPVT-4 wave 4 sum score</b>					
Mean (sd)	49.11 (28.25)	37.37 (27.09)	46.29 (27.84)	52.42 (28.45)	52.94 (27.74)
Median	55	39	52	60	58
<b>PPVT-4 wave 6 sum score</b>					
Mean (sd)	83.92 (22.36)	71.05 (21.68)	80.54 (22.21)	86.29 (23.00)	88.69 (20.75)
Median	84	72	80	85	89

*Note:* NEPS-SC1. Own calculations.

### Clarifications on causal mediation analysis

The casual mediation analysis I am using in the paper is suited for the case of a single exposure ( $SES$ ), whose total effect on an outcome ( $Y$ ), goes through multiple sequential mediators ( $M^P, M^I, M^S$ ). To deal with mediator-outcome confounders affected by exposure, and other confounders, I employ the g-formula approach. Following the diagram in Figure 1 in the paper, this approach relies on the estimation of a series of models: an outcome regression for the language skills of children at a given age ( $Y = f(SES, M, C, C^M)$ ); various mediator regressions of parenting mechanisms ( $M^D = g_D(SES, C, C^M)$  for each of the  $D$  parenting dimensions (P, I, or S) – and more precisely each of the mediators that are being considered); and regressions for mediator-outcome confounders affected by the exposure ( $C^M = w_l(SES, C)$  for each exposure-induced confounder  $l$  affected by exposure  $SES$ ), where  $f$ ,  $g_m$  and  $w_l$  are functions to be estimated. These functions are estimated via generalized linear models, adjusted to each type of dependent variable (i.e., linear, binary, or ordinal). After estimating these models, and based on its predictions, the g-formula is applied to obtain the counterfactual outcomes and compute the respective randomized direct and indirect effects. In the paper, I have outcomes at three time points,  $Y_1$ ,  $Y_2$ , and  $Y_3$ , hence, for later outcomes, all the in-between mediators and exposure-induced confounders, early, middle, and late, are used to compute the counterfactual distribution. In this sense, the hypothetical intervention corresponds to a sustained intervention, not just a one point in time. Therefore, for later outcomes, the mediating mechanisms become more complex, involving early and later mediators. In the models, I include all parenting mediators that have taken place before the measurement of the outcome, and adjust for all observed potential exposure induced confounders in between, as described in the main paper.

# Robustness check employing mother’s educational attainment instead of latent class approach to SES

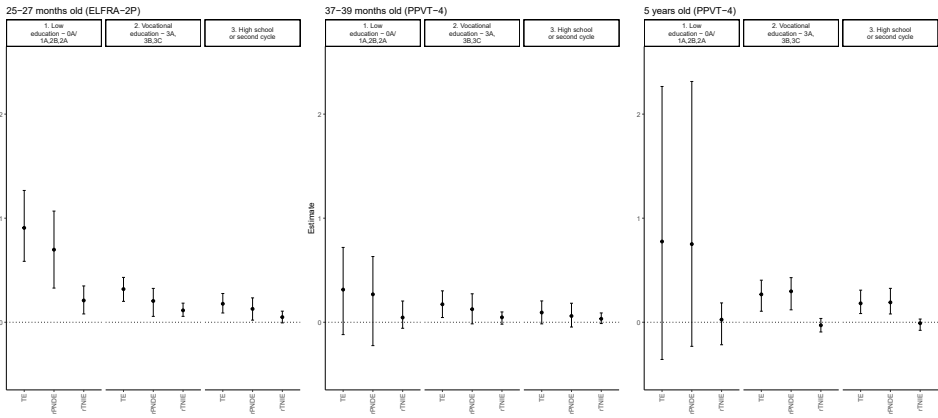


Figure S 1: Interventional/Randomized mediation effects by maternal education (ISCED) on language skills at three time points with respect to mothers with university education

## **Supplementary Materials - Chapter 5**

### **Sample Descriptive Statistics**

Descriptive statistics of samples used in the paper are shown in Table S7 and S8.

### **Latent Class Analysis for the Social Class Structure**

Latent class analysis (LCA) was used to identify homogeneous groups in the data using three widely used socioeconomic characteristics: a) the educational level of mother and father; b) the occupations of mother and father classified by the occupational class structure of Eriksson and Goldthorpe (see Evans 1992); and c) the monthly household adjusted income level as reported by the child's parents. These characteristics are associated to socioeconomic status (SES), an unobservable variable. An inductive or formative model is used to estimate SES, as in latent class analysis following the work of Savage et al. (2013). Although researchers tend to associate an order among the set of latent classes, these classes constitute different categories without an intrinsic order. An assumption in LCA is that the different variables making up the latent classes are assumed to be conditionally independent given that the observations, in this case children, belong to the same class. The number of groups or classes was chosen based on statistical criteria, given that no theoretical number of SES strata is acknowledged. The model with 4 latent classes was chosen as providing the best fit (results not shown). As seen in Table 10, classification in four groups was still interpretable. Low-SES (Class 4) are children whose parents are mostly low educated, have low occupational attainment and are in the low category of household income. The most relevant group of contrast is against children with highly educated parents, high ranking occupations and a high household income (class 2), which serves as the

reference category in all analysis in the paper.

### **Parametric Item Response Theory (PIRT): partial credit model for polytomous item responses**

This section presents the results of the traditional validation framework following the *Standards for Educational and Psychological Testing* (Association, Association, and Measurement in Education 2014). Table S11 shows the frequency distribution of responses to each of the items composing the test, and percentage of missing values. None of the items has low variance or a substantial percentage of missing values. At this stage, all items can be considered for the next analysis. The polychoric correlation shown in Figure S2 suggests that inter-item correlations are on the middle to low range, although some negative correlations were also found. Not all items vary in the same direction, as would be expected from a mathematics test, unless skill in some types of questions is negatively associated to others, which seems implausible. These items were however not excluded at this stage. For reliability, coefficients  $\alpha$ , Guttman's  $\lambda_2$ , and hierarchical  $\omega_h$  are computed (Revelle and Zinbarg 2009). Coefficients Omega hierarchical  $\omega_h = 0.797$ , Cronbach's  $\alpha = 0.874$  and Guttman's  $\lambda_6 = 0.764$  show high values. In addition, inter-item correlations, as well as drop-item reliabilities, are computed to examine internal reliability. Inter-item and drop-item statistics presented in Table S12 indicate sufficient internal reliability of the mathematics scale built out of these 20 items.

An exploration of the underlying structure present in the data is obtained by computing EFA, CFA and structural equation models (SEM). For the EFA, I chose a larger number of hypothesized factors in the mathematics test in order to compare the fit of different solutions. I then used parallel analysis, Very Simple Structure (VSS) criteria and the Velicer MAP criterion to compare solutions of varying complexity. The solution with one factor was

selected and then estimated using CFA, and the fit of the model is assessed through the Normed Fit Index (NFI), Non-Normed Fit Index (NNFI), Comparative Fit Index (CFI), and root mean square error of approximation (RMSEA), all of which are used to assess the global fit of latent variable models. Results of the exploratory factor analysis are shown in Table S13. These suggests, according to the different criteria, that a one factor corresponds to the best solution. This was followed by the estimation of a confirmatory maximum likelihood two-bifactor model (Chalmers 2012). For this I used the M2 statistic; RMSEA; the Standardized Root Mean Square Residual (SRMSR); the Tucker Lewis index (TLI); and again the CFI. As shown in Table S14 and S15, taken together, the items in the mathematics test have appropriate fit indices according to typical CFA and the bi-factor model criteria, substantiating the claim that items are measuring the unidimensional underlying construct of mathematics ability in children.

The partial credit model for polytomous item responses was fitted to the mathematics test data. As shown in Table S16, the model provides estimates of item location and category threshold parameters (Thorpe and Favia 2012). The fit of the model show convergence problems, which are possibly caused by item r14s not fitting the assumptions of the model; the probability of answering 2 out of 3 correct options in this item is never above 50% for some levels of ability. A transformation of this item to a dichotomous one generates the same output as found by Petersen and Gerken (2018). This might be the reason why in Petersen and Gerken (2018) this variable was recoded as dichotomous before fitting the model that was used to compute the NEPS SC1 mathematics ability estimates.

Additionally, these properties can be analyzed following the nonparametric item response theory. In it, unidimensionality claims that manifest responses to items are caused by one single attribute, construct or skill. Unidimensionality is assumed by the three standardized tests. Most IRT

models assume unidimensionality too, and even though multidimensional IRT modes exist, their use is rare. No unique method to assess unidimensionality exists, but in the traditional validation framework the examination of factor loadings and eigenvalues generated by EFA is considered sufficient. The monotone homogeneity model (MHM), however, assess dimensionality of a scale by examining the behavior of a family of scalability coefficients—the  $H_{jk}$ ,  $H_j$  and  $H$  coefficients (Sijtsma and Ark 2016, 145)—as the requirement to conform a scale of weak, medium or strong association is explored. The automatic item selection procedure (AISP) with the genetic algorithm has been shown to perform better in simulation studies to examine structure, though there are some limitations when seeking to discover an underlying structure, as explained in Straat, Van der Ark, and Sijtsma (2013) through a simulation study. The assumption of monotonicity in this MHM model refers to the item step response function (ISRF). A monotonic ISRF refers to  $\mathbb{P}(Y_j \geq y_j | Y^\theta)$  being a non decreasing on the latent attribute  $Y^\theta$  for all  $j$  items. As the construct increases, the probability of correctly answering an item should be higher and likewise more difficult items should require higher values of ability. Number of violations of monotonicity assumption are presented. Local Independence refers Conditioning on the attribute  $Y^\theta$ , items  $j, k$  are independent for all pairs  $(j, k)$ . The indices  $W_1$  and  $W_3$  present items flagged for local independence violations. Invariant Item Ordering: This property does not correspond to the MHM, but to the double homogeneity model. It states that all items are scored in the same order by all individuals responding to the test, at all levels of ability. This was assessed by the  $H^T$  coefficient.

The progression of the mathematics scale by increasing the threshold  $\psi$  as shown in supplementary materials Table S17 revealed that 7 items are unscalable and 2 items belong to another scale. No items were found to

have  $H_j < 0$ , but inter-item scalability coefficients  $H_{jk} < 0$  were present, as well as monotonicity and local independence violations. Finally, Guttman errors, which measure the extent to which items may behave as in a unidimensional series, meaning that harder items are only answered if the easier items have also been correctly answered. On the basis of the outlier score  $G_+$ , which is discussed in Sijtsma and Ark (2016), the distribution of children and number of Guttman errors is presented in a plot for the original  $J$  items and the subset of items conforming to a Mokken scale. Figure 3 presents the distribution of Guttman errors for the full and the MKS scales; these contain a considerable number of Guttman errors, but the MKS shows considerably fewer ones.

## Tables and Figures

### *Tables*

Table S8: Sample Descriptive Statistics Complete Cases in NEPS SC1 5th Wave

Variable	Categories	N	%	ELFRA-2 Productive at Wave 3				PPVT-4 at Wave 4				Mathematics Test at Wave 5			
				Mean	Median	Min.	Max.	Mean	Median	Min.	Max.	Mean	Median	Min.	Max.
Preterm birth	Full-term	1122	93.81	152	161	2	260	48.4	54.0	0	116	0.054	0.048	-3.56	3.19
	Preterm	74	6.19	142	156	14	251	47.1	51.0	2	108	-0.067	-0.016	-3.56	3.19
Gender	Boy	565	47.24	144	155	4	260	48.7	55.0	0	112	-0.042	-0.010	-3.56	3.19
	Girl	631	52.76	158	167	2	260	48.0	53.0	0	116	0.125	0.108	-3.56	3.19
Migration Background	No Migration Background	1101	92.06	155	164	4	260	49.6	56.0	0	116	0.080	0.086	-3.56	3.19
	With migration background	95	7.94	112	121	2	260	33.5	34.0	2	77	-0.342	-0.303	-2.27	1.78
Social Class	SES Class 1	268	22.41	150	156	6	260	49.7	56.0	0	112	0.093	0.089	-2.35	3.19
	SES Class 2	539	45.07	162	172	6	260	51.6	59.0	0	116	0.229	0.229	-2.99	3.19
	SES Class 3	299	25.00	147	158	2	260	45.8	50.0	0	100	-0.088	-0.117	-3.33	2.06
	SES Class 4	90	7.53	109	102	7	260	33.6	35.5	0	83	-0.743	-0.935	-3.56	1.84
Total	-	1196	100.00	152	160	2	260	48.3	54.0	0	116	0.046	0.038	-3.56	3.19

Table S9: Sample Descriptive Statistics Complete Cases by Standardized Test in NEPS SC1

Variable	Categories	ELFRA-2 Productive at Wave 3						PPVT-4 at Wave 4						Mathematics Test at Wave 5					
		N	%	Mean	Median	Min.	Max.	N	%	Mean	Median	Min.	Max.	N	%	Mean	Median	Min.	Max.
Preterm birth	Full-term	1997	93.8	141.0	151	0	260	1580	94.1	46.5	52.0	0	121	1729	94.28	-0.001	0.012	-3.56	3.19
	Preterm	132	6.2	124.1	133	1	251	99	5.9	42.9	47.0	0	108	105	5.72	-0.196	-0.137	-3.56	3.19
Gender	Boy	1061	49.8	131.1	144	0	260	820	48.8	45.8	52.0	0	121	905	49.35	-0.092	-0.071	-3.56	3.19
	Girl	1068	50.2	148.8	158	1	260	859	51.2	46.8	52.0	0	116	929	50.65	0.066	0.094	-3.56	3.19
Migration Background	No Migration Background	1892	88.9	146.1	157	0	260	1503	89.5	48.1	54.0	0	121	1634	89.09	0.041	0.053	-3.56	3.19
	With migration background	237	11.1	91.0	82	0	260	176	10.5	30.9	31.5	0	86	200	10.90	-0.442	-0.390	-3.31	2.53
Social Class	SES Class 1	459	21.6	145.2	153	5	260	365	21.7	47.8	54.0	0	112	408	22.25	0.066	0.065	-2.66	3.19
	SES Class 2	851	40.0	156.8	167	2	260	700	41.7	51.0	58.0	0	117	759	41.38	0.209	0.207	-2.99	3.19
	SES Class 3	562	26.4	130.0	140	0	260	432	25.7	43.4	47.0	0	121	487	26.55	-0.152	-0.137	-3.33	2.53
	SES Class 4	257	12.1	96.9	88	1	260	182	10.8	32.3	34.0	0	98	180	9.81	-0.741	-0.863	-3.56	1.84
Total	-	2129	100.0	140.0	151	0	260	1679	100.0	46.3	52.0	0	121	1834	100.00	-0.012	0.002	-3.56	3.19

Table S10: Latent Class Analysis for Social Class Structure in NEPS SC1 Cohort at Wave 1

Variable	Categories	Latent Classes			
		Class 1	Class 2	Class 3	Class 4
Mother's educational level	No degree or vocational/voluntary degree (Haupt-, Real-, Volksschulabschluss)	2.32	0.00	2.79	51.52
	Technical/applied or Civil Servant	33.94	6.11	48.97	39.46
	Technical Degree (Fachhochschulreife)	47.57	10.30	34.73	8.86
	University Education	16.17	83.59	13.50	0.17
Father's educational level	No degree or vocational/voluntary degree (Haupt-, Real-, Volksschulabschluss)	1.36	0.31	6.61	45.73
	Technical/applied or Civil Servant	16.51	8.91	71.24	46.52
	Technical Degree (Fachhochschulreife)	39.31	8.20	22.16	5.21
	University Education	42.82	82.58	0.00	2.55
Mother's EGP occupational class	I and II	44.40	93.15	48.64	9.45
	IIIa and IIIb	45.69	5.15	40.91	50.19
	IVa, IVb and IVc	3.59	1.43	1.64	2.31
	V and VI	2.84	0.27	6.04	9.43
Father's EGP occupational class	VIIa and VIIb	3.48	0.00	2.78	28.63
	I and II	78.04	95.48	23.20	9.65
	IIIa and IIIb	11.41	2.08	16.89	11.56
	IVa, IVb and IVc	7.29	1.90	4.64	4.94
Household adjusted monthly income in 2012 EUR	V and VI	0.00	0.41	36.95	23.86
	VIIa and VIIb	3.26	0.13	18.32	50.00
	(0,1160]	15.02	7.36	28.94	80.44
	(1160,1620]	24.80	13.69	44.61	17.12
	(1620,2019]	34.11	29.49	23.47	1.62
	(2190,16200]	26.07	49.46	2.99	0.82

## Figures

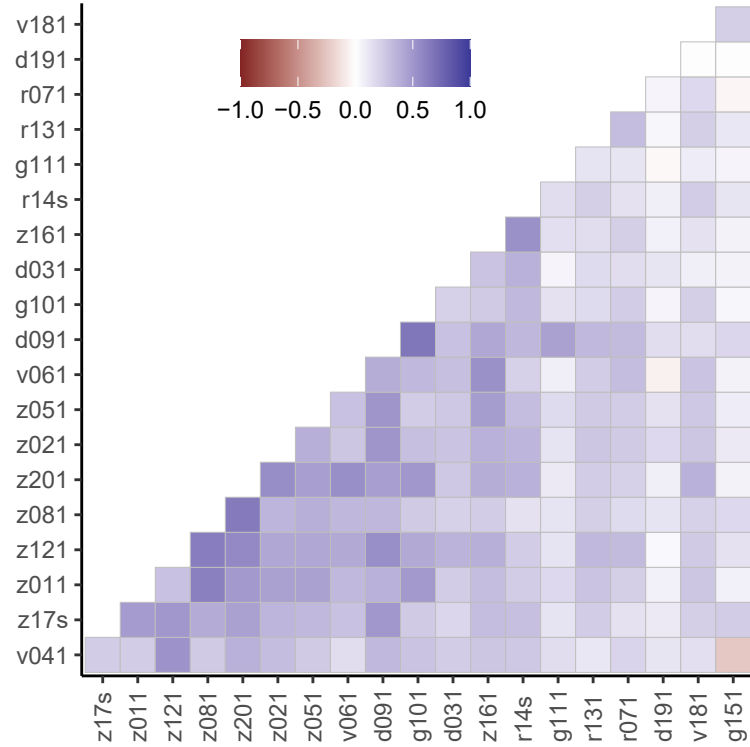


Figure S 2: Polychoric Correlation for Mathematics Test Items

Table S11: Mathematics Test Items

	Missing		Incorrect or Wrong		Correct: Right answer   1 out of 3		Correct: 2 out of 3		Correct: 3 out of 3	
	n	%	n	%	n	%	n	%	n	%
z17s	0	0.00	575	27.38	440	21.72	615	30.5	410	20.4
z021	0	0.00	948	45.94	1092	54.06	-	-	-	-
v181	0	0.00	1350	66.05	690	33.95	-	-	-	-
z161	3	0.15	649	31.24	1388	68.76	-	-	-	-
r14s	6	0.29	105	4.76	293	14.44	302	14.9	1334	65.9
d191	8	0.39	1040	51.10	992	48.90	-	-	-	-
z051	10	0.49	1383	68.00	647	32.00	-	-	-	-
g151	10	0.49	578	28.44	1452	71.56	-	-	-	-
r131	12	0.59	1369	67.35	659	32.65	-	-	-	-
g111	15	0.74	1749	86.36	276	13.64	-	-	-	-
z121	16	0.78	292	14.14	1732	85.86	-	-	-	-
v041	21	1.03	1358	67.05	661	32.95	-	-	-	-
z081	21	1.03	1893	93.73	126	6.27	-	-	-	-
d091	23	1.13	199	9.83	1818	90.17	-	-	-	-
z201	26	1.27	1236	61.33	778	38.67	-	-	-	-
g101	29	1.42	394	19.56	1617	80.44	-	-	-	-
z011	29	1.42	1374	68.20	637	31.80	-	-	-	-
r071	43	2.11	991	49.55	1006	50.45	-	-	-	-
d031	31	1.52	1422	70.71	587	29.29	-	-	-	-
v061	32	1.57	1138	56.67	870	43.33	-	-	-	-

Table S12: Internal Reliability Coefficients for Mathematics Test Items

	Cronbach's Alpha	Guttman's Lambda 6	Average interitem cor.	Median interitem cor.	Cor. with score (corrected)	Drop item cor.
z17s	0.865	0.919	0.252	0.240	0.614	0.552
z021	0.863	0.925	0.250	0.238	0.654	0.597
v181	0.871	0.975	0.263	0.249	0.451	0.373
z161	0.865	0.975	0.253	0.238	0.605	0.542
r14s	0.868	0.948	0.256	0.244	0.549	0.479
d191	0.880	0.976	0.278	0.256	0.214	0.124
z051	0.864	0.928	0.251	0.238	0.635	0.575
g151	0.879	0.990	0.277	0.256	0.231	0.141
r131	0.871	0.953	0.262	0.249	0.468	0.392
g111	0.876	0.944	0.271	0.254	0.325	0.239
z121	0.860	0.913	0.245	0.238	0.728	0.680
v041	0.870	0.945	0.261	0.243	0.479	0.404
z081	0.864	0.914	0.251	0.243	0.634	0.574
d091	0.859	0.906	0.243	0.238	0.765	0.722
z201	0.860	0.926	0.244	0.238	0.750	0.705
g101	0.866	0.905	0.254	0.238	0.590	0.524
z011	0.863	0.903	0.250	0.238	0.655	0.597
r071	0.872	0.983	0.263	0.251	0.441	0.363
d031	0.870	0.971	0.261	0.246	0.475	0.399
v061	0.866	0.964	0.254	0.238	0.581	0.515

Table S13: Criteria for Number of Factors in Mathematics Test Items

Factors	VSS 1	VSS 2	MAP	Parallel FA
1	0.498	0.000	0.005	3.130
2	0.364	0.413	0.007	0.435
3	0.319	0.413	0.011	0.179
4	0.251	0.359	0.015	0.157
5	0.246	0.343	0.021	0.128
6	0.250	0.325	0.027	0.084

Table S14: Confirmatory Factor Analysis for Mathematics Test Items

Chi-Square	d.f.	P-value	NFI	NNFI	CFI	RMSEA
322	170	1.64e-11	0.972	0.985	0.987	0.021

Table S15: Full-Information Item Bi-factor and Two-Tier Analysis for Mathematics Test Items

M2	d.f.	P-value	RMSEA	SRMSR	TLI	CFI
237	146	2.67e-06	0.018	0.027	0.984	0.988

Table S16: Graded Response Model Estimates for Mathematics Test Items

	Threshold: 1 vs. 0   1/3	Threshold: 2/3	Threshold: 3/3	Discrimination
z17s	-0.128	-0.44	1.1	0.71
z021	-0.153	-	-	1.42
v181	0.981	-	-	0.77
z161	-0.889	-	-	1.09
r14s	-2.865	-0.48	-2.6	0.56
d191	0.215	-	-	0.20
z051	0.821	-	-	1.16
g151	-3.386	-	-	0.28
r131	1.066	-	-	0.77
g111	4.369	-	-	0.44
z121	-1.455	-	-	1.90
v041	0.980	-	-	0.83
z081	2.089	-	-	1.91
d091	-1.937	-	-	1.57
z201	0.361	-	-	2.37
g101	-1.627	-	-	1.05
z011	0.675	-	-	1.69
r071	-0.028	-	-	0.67
d031	1.272	-	-	0.78
v061	0.299	-	-	1.15

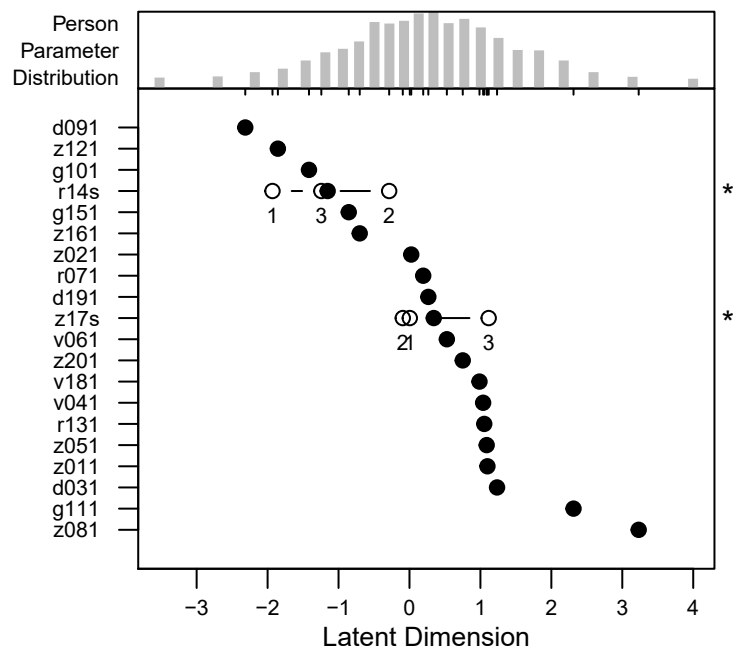


Figure S 3: Person and Item Fit Plot from Graded Response Model in Mathematics Test Items

Table S17: AISP Genetic Algorithm for Mathematics Test Items

	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55
z17s	1	1	1	1	1	1	1	0	1	1	0	0
z021	1	1	1	1	1	1	1	1	0	0	0	0
v181	1	1	1	1	1	0	0	2	0	0	0	0
z161	1	1	1	1	1	1	1	1	0	0	0	0
r14s	1	1	1	1	1	1	2	3	3	0	0	0
d191	2	2	0	0	0	0	0	0	0	0	0	0
z051	1	1	1	1	1	1	1	1	3	0	0	0
g151	1	1	0	2	0	0	0	0	0	0	0	0
r131	1	1	1	1	1	0	2	0	0	0	0	0
g111	1	1	1	1	0	0	0	0	0	0	0	0
z121	1	1	1	1	1	1	1	1	1	1	1	1
v041	2	2	1	1	1	1	0	3	0	0	0	0
z081	1	1	1	2	1	1	1	1	1	1	1	1
d091	1	1	1	1	1	1	1	1	1	1	1	2
z201	1	1	1	1	1	1	1	1	1	1	1	0
g101	1	1	1	1	1	1	1	2	2	0	0	0
z011	1	1	1	1	1	1	1	1	1	1	1	1
r071	1	1	1	1	0	0	0	0	0	0	0	0
d031	1	1	1	1	1	0	0	0	0	0	0	0
v061	1	1	1	1	1	1	1	0	2	0	0	2

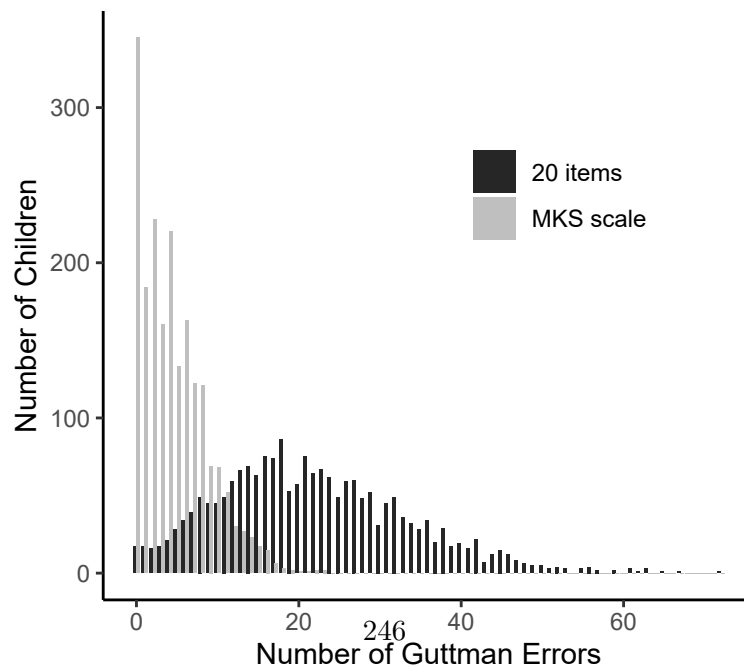


Figure S 4: Number of Gutman Errors and Number of Children in the Mathematics Test Items and their Mokken Subscale

Table S18: Mathematics Items Flagged as DIF by Generalized Mantel-Haezel Test

Items	Preterm		Girls		Mig. Background		Low-SES	
	gMH	P-value	gMH	P-value	gMH	P-value	gMH	P-value
z17s	5.51	0.02	18.46	1.73e-05	5.96	0.01	5.96	0.01
z021	0.35	0.55	0.72	0.40	6.03	0.01	0.01	0.93
v181	0.85	0.36	7.28	6.97e-03	0.03	0.87	0.23	0.63
z161	0.00	0.96	0.00	0.95	4.92	0.03	0.57	0.45
r14s	14.78	1.21e-04	40.33	2.15e-10	65.12	6.66e-16	5.67	0.02
d191	5.35	0.02	2.14	0.14	1.21	0.27	1.85	0.17
z051	0.04	0.84	0.09	0.77	4.09	0.04	1.09	0.30
g151	0.07	0.79	0.14	0.71	2.07	0.15	3.28	0.07
r131	1.07	0.30	0.21	0.65	0.03	0.87	1.76	0.19
g111	3.87	0.05	6.18	0.01	3.13	0.08	1.31	0.25
z121	4.30	0.04	1.32	0.25	0.85	0.36	1.76	0.18
v041	0.31	0.58	0.20	0.65	0.03	0.86	1.00	0.32
z081	0.01	0.94	12.64	3.77e-04	0.81	0.37	0.12	0.73
d091	0.98	0.32	15.51	8.19e-05	0.05	0.82	0.26	0.61
z201	0.68	0.41	0.00	0.95	0.27	0.61	0.05	0.82
g101	1.51	0.22	0.02	0.89	2.41	0.12	3.74	0.05
z011	0.58	0.45	1.24	0.27	0.00	0.95	0.46	0.50
r071	0.59	0.44	0.97	0.32	0.04	0.83	0.19	0.67
d031	0.19	0.67	0.77	0.38	0.46	0.50	2.69	0.10
v061	0.34	0.56	28.27	1.05e-07	6.70	9.66e-03	13.67	2.18e-04

## Bibliography

- Abbott, Andrew. 1988. "Transcending General Linear Reality." *Sociological Theory*, 169–86.
- . 2016. *Processual Sociology*. University of Chicago Press.
- Agresti, Alan. 2012. *Analysis of Ordinal Categorical Data*. New Jersey: John Wiley & Sons.
- Aisenbrey, Silke, and Anette Fasang. 2017. "The Interplay of Work and Family Trajectories over the Life Course: Germany and the United States in Comparison." *American Journal of Sociology* 122 (5): 1448–84.
- Albelda, Randy, Susan Himmelweit, and Jane Humphries. 2004. "The Dilemmas of Lone Motherhood: Key Issues for Feminist Economics." *Feminist Economics* 10 (2): 1–7.
- Amato, Paul R. 2001. "Children of Divorce in the 1990s: An Update of the Amato and Keith (1991) Meta-Analysis." *Journal of Family Psychology* 15 (3): 355.

- . 2003. “Reconciling Divergent Perspectives: Judith Wallerstein, Quantitative Family Research, and Children of Divorce.” *Family Relations* 52 (4): 332–39.
- . 2010. “Research on Divorce: Continuing Trends and New Developments.” *Journal of Marriage and Family* 72 (3): 650–66. <https://doi.org/10.1111/j.1741-3737.2010.00723.x>.
- Amato, Paul R, and Bruce Keith. 1991. “Parental Divorce and the Well-Being of Children: A Meta-Analysis.” [https://slatestarcodex.com/Stuff/divorce\\_paper.pdf](https://slatestarcodex.com/Stuff/divorce_paper.pdf); American Psychological Association.
- Amato, Paul R, and Sarah E Patterson. 2017. “The Intergenerational Transmission of Union Instability in Early Adulthood.” *Journal of Marriage and Family* 79 (3): 723–38.
- Andersson, Gunnar. 2020. “A Review of Policies and Practices Related to the "Highest-Low" Fertility of Sweden: A 2020 Update.”
- Association, American Educational Research, American Psychological Association, and National Council on Measurement in Education. 2014. American Psychological Association.
- Attig, Manja, and Sabine Weinert. 2020. “What Impacts Early Language Skills? Effects of Social Disparities and Different Process Characteristics of the Home Learning Environment in the First 2 Years.” *Frontiers in Psychology* 11: 3218.
- Austin, Peter C, Nathaniel Jembere, and Maria Chiu. 2018. “Propensity Score Matching and Complex Surveys.” *Statistical Methods in Medical Research* 27 (4): 1240–57.
- Ayoub, Catherine, Claire D Vallotton, and Ann M Mastergeorge. 2011. “Developmental Pathways to Integrated Social Skills: The Roles of Parenting and Early Intervention.” *Child Development* 82 (2): 583–600.
- Bailey, Drew, Greg J Duncan, Candice L Odgers, and Winnie Yu. 2017. “Persistence and Fadeout in the Impacts of Child and Adolescent Interventions.” *Journal of Research on Educational Effectiveness* 10 (1): 7–39.
- Baird, Jo-Anne, David Andrich, Therese N. Hopfenbeck, and Gordon Stobart. 2017. “Assessment and Learning: Fields Apart?” *Assessment in Education: Principles, Policy & Practice* 24 (3): 317–50. <https://doi.org/10.1080/0969594X.2017.1319337>.
- Ballou, Dale. 2009. “Test Scaling and Value-Added Measurement.” *Education Finance and Policy* 4 (4): 351–83. <https://doi.org/10.1162/edfp.2009.4.4.351>.
- Banerjee, Pallavi Amitava. 2016. “A Systematic Review of Factors Linked to Poor Academic Performance of Disadvantaged Students in Science and Maths in Schools.” *Cogent Education* 3: 17. <https://doi.org/10.1080/2331186X.2016.1178441>.
- Baumrind, Diana. 2005. “Patterns of Parental Authority and Adolescent

- Autonomy.” *New Directions for Child and Adolescent Development* 2005 (108): 61–69.
- Bear, George G, and Kathleen M Minke. 2006. *Children’s Needs III: Development, Prevention, and Intervention*. National Association of School Psychologists.
- Becker, Birgit. 2011. “Cognitive and Language Skills of Turkish Children in Germany: A Comparison of the Second and Third Generation and Mixed Generational Groups.” *International Migration Review* 45 (2): 426–59.
- Becker, Gary S. 1973. “A Theory of Marriage: Part i.” *Journal of Political Economy* 81 (4): 813–46.
- . 1991. *A Treatise on the Family: Enlarged Edition*. Harvard university press.
- . 2009. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago press.
- Berendes, Karin, Sabine Weinert, Stefan Zimmermann, and Cordula Artelt. 2013. “Assessing Language Indicators Across the Lifespan Within the German National Educational Panel Study (NEPS).” *Journal for Educational Research Online* 5 (2): 15–49.
- Berge, Jerica M, Melanie Wall, Katherine W Bauer, and Dianne Neumark-Sztainer. 2010. “Parenting Characteristics in the Home Environment and Adolescent Overweight: A Latent Class Analysis.” *Obesity* 18 (4): 818–25.
- Berger, Lawrence M, Christina Paxson, and Jane Waldfogel. 2009. “Income and Child Development.” *Children and Youth Services Review* 31 (9): 978–89.
- Bernardi, Fabrizio, and Diederik Boertien. 2017. “Non-Intact Families and Diverging Educational Destinies: A Decomposition Analysis for Germany, Italy, the United Kingdom and the United States.” *Social Science Research* 63: 181–91.  
<https://doi.org/10.1016/j.ssresearch.2016.09.004>.
- Bernardi, Fabrizio, Juho Härkönen, Diederik Boertien, Linu Andersson, Kim Bastiaits, and Dimitri Mortelmans. 2013. “Effects of Family Forms and Dynamics on Children’s Well-Being and Life Chances: Literature Review.”
- Bernardi, Laura, Johannes Huinink, and Richard A Settersten Jr. 2019. “The Life Course Cube: A Tool for Studying Lives.” *Advances in Life Course Research* 41: 100258.
- Bernstein, Basil. 1964. “Elaborated and Restricted Codes: Their Social Origins and Some Consequences.” *American Anthropologist* 66 (6): 55–69.
- Bhrolchain, Maire Ni. 2001. “‘Divorce Effects’ and Causality in the Social Sciences.” *European Sociological Review* 17 (1): 33–57.
- Bianchi, Suzanne M. 2014. “A Demographic Perspective on Family Change.” *Journal of Family Theory & Review* 6 (1): 35–44.
- Bihagen, Erik, and Paul Lambert. 2018. “Can Class and Status Really Be

- Disentangled?" *Research in Social Stratification and Mobility* 58: 1–10.
- Bloome, Deirdre. 2017. "Childhood Family Structure and Intergenerational Income Mobility in the United States." *Demography* 54 (2): 541–69. <https://doi.org/10.1007/s13524-017-0564-4>.
- Blossfeld, Hans-Peter, Hans-Günther Roßbach, and Jutta von Maurice. 2011. "Education as a Lifelong Process - the German National Educational Panel Study (NEPS)." *Zeitschrift für Erziehungswissenschaft* 14 (2): 19–34. <https://doi.org/10.1007/s11618-011-0179-2>.
- Bond, Tymothy N., and Kevin Lang. 2013. "The Black-White Education Scaled Test-Score Gap in Grades k-7." *NBER Working Paper*, no. 19243: 39. <https://doi.org/10.3386/w19243>.
- Bongers, Ilja L, Hans M Koot, Jan Van der Ende, and Frank C Verhulst. 2003. "The Normative Development of Child and Adolescent Problem Behavior." *Journal of Abnormal Psychology* 112 (2): 179.
- Bornstein, Marc H. 2012. "Cultural Approaches to Parenting." *Parenting* 12 (2-3): 212–21.
- Borsboom, Denny. 2006. "The Attack of the Psychometricians." *Psychometrika* 71 (3): 425–40. <https://doi.org/10.1007/s11336-006-1447-6>.
- Borsboom, Denny, Gideon J. Mellenbergh, and Jaap van Heerden. 2004. "The Concept of Validity." *Psychological Review* 111 (4): 1061–71. <https://doi.org/10.1037/0033-295X.111.4.1061>.
- Borsboom, Denny, and Lisa D. Wijsen. 2017. "Psychology's Atomic Bomb." *Assessment in Education: Principles, Policy & Practice* 24 (3): 440–46. <https://doi.org/10.1080/0969594X.2017.1333084>.
- Bourdieu, Pierre. 1984. "La Distinction. Paris: Editions de Minuit, 1979." *Trans. By Richard Nice as Distinction: A Social Critique of the Judgment of Taste. Cambridge, MA: Harvard UP.*
- . 1991. *Language and Symbolic Power*. Harvard University Press.
- . 1993. "A Propos de La Famille Comme Catégorie réalisée." *Actes de La Recherche En Sciences Sociales*, no. 100: 32–36.
- Bourdieu, Pierre, and Jean-Claude Passeron. 1964. *Les héritiers. Les étudiants Et La Culture*. Paris: Editions de Minuit.
- Bradley, Robert H, and Robert F Corwyn. 2002. "Socioeconomic Status and Child Development." *Annual Review of Psychology* 53 (1): 371–99.
- Brady, David. 2019. "Theories of the Causes of Poverty." *Annual Review of Sociology* 45: 155–75.
- Brand, Jennie E, Ravaris Moore, Xi Song, and Yu Xie. 2019. "Why Does Parental Divorce Lower Children's Educational Attainment? A Causal Mediation Analysis." *Sociological Science* 6: 264–92. <https://doi.org/10.15195/v6.a11>.
- Brandone, Amanda C, Sara J Salkind, Roberta Michnick Golinkoff, and Kathy Hirsh-Pasek. 2006. "Language Development."
- Briggs, Derek C. 2013. "Measuring Growth with Vertical Scales." *Journal of*

- Educational Measurement* 50 (2): 204–26.  
<https://doi.org/10.1111/jedm.12011>.
- Bronfenbrenner, Urie, and Pamela A Morris. 2007. “The Bioecological Model of Human Development.” *Handbook of Child Psychology* 1.
- Brown, Susan L, Wendy D Manning, and J Bart Stykes. 2015. “Family Structure and Child Well-Being: Integrating Family Complexity.” *Journal of Marriage and Family* 77 (1): 177–90.  
<https://doi.org/10.1111/jomf.12145>.
- Browning, Martin, Pierre-André Chiappori, and Yoram Weiss. 2014. *Economics of the Family*. Cambridge University Press.
- Burgess, Ernest W. 1950. “Predictive Methods and Family Stability.” *The ANNALS of the American Academy of Political and Social Science* 272 (1): 47–52.
- Burgette, Lane F, and Jerome P Reiter. 2010. “Multiple Imputation for Missing Data via Sequential Regression Trees.” *American Journal of Epidemiology* 172 (9): 1070–76. <https://doi.org/10.1093/aje/kwq260>.
- Cahill, Sean. 2005. “Welfare Moms and the Two Grooms: The Concurrent Promotion and Restriction of Marriage in US Public Policy.” *Sexualities* 8 (2): 169–87.
- Cameron, A Colin, and Pravin K Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge university press.
- Carnegie, Nicole Bohme, and James Wu. 2019. “Variable Selection and Parameter Tuning for BART Modeling in the Fragile Families Challenge.” *Socius* 5: 2378023119825886.
- Cavanagh, Shannon E, and Paula Fomby. 2019. “Family Instability in the Lives of American Children.” *Annual Review of Sociology* 45.  
<https://doi.org/10.1146/annurev-soc-073018-022633>.
- Cavanagh, Shannon E, and Aletha C Huston. 2006. “Family Instability and Children’s Early Problem Behavior.” *Social Forces* 85 (1): 551–81.
- . 2008. “The Timing of Family Instability and Children’s Social Development.” *Journal of Marriage and Family* 70 (5): 1258–70.
- Census Bureau, U. S. 2019. “Historical Living Arrangements of Children: 1960 to Present.” 2019.  
<https://www.census.gov/library/visualizations/time-series/demo/children-historical-time-series.html>.
- Chalmers, R. Philip. 2012. “Mirt: A Multidimensional Item Response Theory Package for the r Environment.” *Journal of Statistical Software* 48 (6): 1–29. <https://doi.org/10.18637/jss.v048.i06>.
- Chan, Eric K. H. 2014. “Standards and Guidelines for Validation Practices: Development and Evaluation of Measurement Instruments.” In *Validity and Validation in Social, Behavioral, and Health Sciences*, edited by B. D. Zumbo, Eric K. H., and E. K. H. Chan, 9–24. Springer.
- Chan, Tak Wing, and Anita Koo. 2011. “Parenting Style and Youth Outcomes in the UK.” *European Sociological Review* 27 (3): 385–99.

- Cherlin, Andrew J. 2004. "The Deinstitutionalization of American Marriage." *Journal of Marriage and Family* 66 (4): 848–61.
- . 2010. "Demographic Trends in the United States: A Review of Research in the 2000s." *Journal of Marriage and Family* 72 (3): 403–19.
- Chipman, Hugh A, Edward I George, Robert E McCulloch, et al. 2010. "BART: Bayesian Additive Regression Trees." *The Annals of Applied Statistics* 4 (1): 266–98.
- Cina, Annette, Guy Bodenmann, Kurt Hahlweg, Thomas Dirscherl, and Matthew R Sanders. 2006. "Triple p (Positive Parenting Program): Theoretischer Und Empirischer Hintergrund Und Erste Erfahrungen Im Deutschsprachigen Raum." *Zeitschrift für Familienforschung* 18 (1): 66–88.
- Cobb-Clark, Deborah A, Nicolas Salamanca, and Anna Zhu. 2019. "Parenting Style as an Investment in Human Development." *Journal of Population Economics* 32 (4): 1315–52.
- Cohen, Philip N. 2002. "Extended Households at Work: Living Arrangements and Inequality in Single Mothers' Employment." In *Sociological Forum*, 17:445–63. 3. Springer.  
<https://doi.org/10.1023/A:1019631107686>.
- Cohen, Philip N, and Joanna R Pepin. 2018. "Unequal Marriage Markets: Sex Ratios and First Marriage Among Black and White Women." *Socius* 4: 2378023118791084. <https://doi.org/10.1177/2378023118791084>.
- Collins, Patricia Hill. 2009. *Black Feminist Thought*. Routledge.
- Coltrane, Scott, and Michele Adams. 2003. "The Social Construction of the Divorce "Problem": Morality, Child Victims, and the Politics of Gender." *Family Relations* 52 (4): 363–72.
- Conger, Rand D, Katherine J Conger, and Monica J Martin. 2010. "Socioeconomic Status, Family Processes, and Individual Development." *Journal of Marriage and Family* 72 (3): 685–704.
- Conrad-Hiebner, Aislinn, and Elizabeth Byram. 2020. "The Temporal Impact of Economic Insecurity on Child Maltreatment: A Systematic Review." *Trauma, Violence, & Abuse* 21 (1): 157–78.
- Coontz, Stephanie. 2016. *The Way We Never Were: American Families and the Nostalgia Trap*. Hachette UK.
- Cooper, Marianne, and Allison J Pugh. 2020. "Families Across the Income Spectrum: A Decade in Review." *Journal of Marriage and Family* 82 (1): 272–99.
- Corak, Miles. 2001. "Death and Divorce: The Long-Term Consequences of Parental Loss on Adolescents." *Journal of Labor Economics* 19 (3): 682–715.
- Cote, Anita T, Kevin C Harris, Constadina Panagiotopoulos, George GS Sandor, and Angela M Devlin. 2013. "Childhood Obesity and Cardiovascular Dysfunction." *Journal of the American College of Cardiology* 62 (15): 1309–19.

- Crosnoe, Robert, Kate Chambers Prickett, Chelsea Smith, and Shannon Cavanagh. 2014. "Changes in Young Children's Family Structures and Child Care Arrangements." *Demography* 51 (2): 459–83.  
<https://doi.org/10.1007/s13524-013-0258-5>.
- Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3): 883–931.  
<https://doi.org/10.3982/ECTA6551>.
- Dahl, Ronald E, Nicholas B Allen, Linda Wilbrecht, and Ahna Ballonoff Suleiman. 2018. "Importance of Investing in Adolescence from a Developmental Science Perspective." *Nature* 554 (7693): 441–50.
- Dalziel, James. 1998. "Using Marks to Assess Student Performance, Some Problems and Alternatives." *Assessment & Evaluation in Higher Education* 23 (4): 351–66. <https://doi.org/10.1080/0260293980230403>.
- Damaske, Sarah, Jenifer L Bratter, and Adrienne Frech. 2017. "Single Mother Families and Employment, Race, and Poverty in Changing Economic Times." *Social Science Research* 62: 120–33.  
<https://doi.org/10.1016/j.ssresearch.2016.08.008>.
- Darling, Nancy, and Laurence Steinberg. 1993. "Parenting Style as Context: An Integrative Model." *Psychological Bulletin* 113 (3): 487.
- De Stavola, Bianca L, Moritz Herle, and Andrew Pickles. 2022. "Framing Causal Questions in Life Course Epidemiology." *Annual Review of Statistics and Its Application* 9: 223–48.
- De Vaus, David, Matthew Gray, Lixia Qu, and David Stanton. 2017. "The Economic Consequences of Divorce in Six OECD Countries." *Australian Journal of Social Issues* 52 (2): 180–99.
- Dehaene-Lambertz, G., and E. S. Spelke. 2015. "The Infancy of the Human Brain." *Neuron* 88 (1): 93–109.  
<https://doi.org/10.1016/j.neuron.2015.09.026>.
- Dermott, Esther, and Marco Pomati. 2016. "'Good'parenting Practices: How Important Are Poverty, Education and Time Pressure?" *Sociology* 50 (1): 125–42.
- Desmond, Matthew, and Kristin L Perkins. 2016. "Housing and Household Instability." *Urban Affairs Review* 52 (3): 421–36.  
<https://doi.org/10.1177/1078087415589192>.
- Devenish, Bethany, Marilyn Hooley, and David Mellor. 2017. "The Pathways Between Socioeconomic Status and Adolescent Outcomes: A Systematic Review." *American Journal of Community Psychology* 59 (1-2): 219–38.
- Dimitrov, Dimitar M. 2016. "An Approach to Scoring and Equating Tests with Binary Items: Piloting with Large-Scale Assessments." *Educational and Psychological Measurement* 76 (6): 954–75.  
<https://doi.org/10.1177/0013164416631100>.
- DiPrete, Thomas A, and Gregory M Eirich. 2006. "Cumulative Advantage

- as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments.” *Annu. Rev. Sociol.* 32: 271–97.
- Doepke, Matthias, Giuseppe Sorrenti, and Fabrizio Zilibotti. 2019. “The Economics of Parenting.” *Annual Review of Economics* 11: 55–84.
- Doepke, Matthias, and Fabrizio Zilibotti. 2017. “Parenting with Style: Altruism and Paternalism in Intergenerational Preference Transmission.” *Econometrica* 85 (5): 1331–71.
- Dollaghan, Christine A, Thomas F Campbell, Jack L Paradise, Heidi M Feldman, Janine E Janosky, Dayna N Pitcairn, and Marcia Kurs-Lasky. 1999. “Maternal Education and Measures of Early Speech and Language.” *Journal of Speech, Language, and Hearing Research* 42 (6): 1432–43.
- Domingue, Ben. 2014. “Evaluating the Equal-Interval Hypothesis with Test Score Scales.” *Psychometrika* 79 (1): 1–19.  
<https://doi.org/10.1007/s11336-013-9342-4>.
- Dorie, Vincent, Jennifer Hill, Uri Shalit, Marc Scott, Dan Cervone, et al. 2019. “Automated Versus Do-It-Yourself Methods for Causal Inference: Lessons Learned from a Data Analysis Competition.” *Statistical Science* 34 (1): 43–68.
- Downey, Douglas B, and Dennis J Condron. 2016. “Fifty Years Since the Coleman Report: Rethinking the Relationship Between Schools and Inequality.” *Sociology of Education* 89 (3): 207–20.  
<https://doi.org/10.1177/0038040716651676>.
- Dräger, Jascha. 2022. “The Role of Parental Wealth in Children’s Educational Pathways in Germany.” *European Sociological Review* 38 (1): 18–36.
- Du Bois, William Edward Burghardt. 1899. *The Philadelphia Negro*. Cosimo, Inc.
- DuGoff, Eva H, Megan Schuler, and Elizabeth A Stuart. 2014. “Generalizing Observational Study Results: Applying Propensity Score Methods to Complex Surveys.” *Health Services Research* 49 (1): 284–303.
- Duncan, Greg J, W Jean Yeung, Jeanne Brooks-Gunn, and Judith R Smith. 1998. “How Much Does Childhood Poverty Affect the Life Chances of Children?” *American Sociological Review*, 406–23.
- Duncan, Greg J., Kathleen M. ZiolGuest, and Ariel Kalil. 2010. “Early-Childhood Poverty and Adult Attainment, Behavior, and Health.” *Child Development* 81 (1): 306–25.  
<https://doi.org/10.1111/j.1467-8624.2009.01396.x>.
- Edin, Kathryn, and Maria Kefalas. 2005. “Unmarried with Children.” *Contexts* 4 (2): 16–22.
- Edin, Kathryn, and Rebecca Joyce Kissane. 2010. “Poverty and the American Family: A Decade in Review.” *Journal of Marriage and Family* 72 (3): 460–79. <https://doi.org/10.1111/j.1741-3737.2010.00713.x>.
- Elder, Glen H. 2018. *Children of the Great Depression: Social Change in Life Experience*. Routledge.

- Elder Jr, Glen H, Michael J Shanahan, and Julia A Jennings. 2015. "Human Development in Time and Place." *Handbook of Child Psychology and Developmental Science*, 1–49.
- Elliott, Sinikka, and Sarah Bowen. 2018. "Defending Motherhood: Morality, Responsibility, and Double Binds in Feeding Children." *Journal of Marriage and Family* 80 (2): 499–520.
- Engels, Friedrich, and Lewis Henry Morgan. 1884. *The Origin of the Family, Private Property and the State*. Foreign Languages Publishing House Moscow.
- Erola, Jani, and Marika Jalovaara. 2017. "The Replaceable: The Inheritance of Paternal and Maternal Socioeconomic Statuses in Non-Standard Families." *Social Forces* 95 (3): 971–95.  
<https://doi.org/10.1093/sf/sow089>.
- Erola, Jani, and Elina Kilpi-Jakonen. 2017. "Compensation and Other Forms of Accumulation in Intergenerational Social Inequality." In *Social Inequality Across the Generations*. Edward Elgar Publishing.
- Esping-Andersen, Gøsta. 2016. *Families in the 21st Century*. SNS Förlag Stockholm.
- Evans, Geoffrey. 1992. "Testing the Validity of the Goldthorpe Class Schema." *European Sociological Review* 8 (3): 211–32.  
<https://doi.org/10.1093/oxfordjournals.esr.a036638>.
- Farah, Martha J. 2017. "The Neuroscience of Socioeconomic Status: Correlates, Causes, and Consequences." *Neuron* 96 (1): 56–71.  
<https://doi.org/10.1016/j.neuron.2017.08.034>.
- Fasang, Anette Eva, and Silke Aisenbrey. 2021. "Uncovering Social Stratification: Intersectional Inequalities in Work and Family Life Courses by Gender and Race." *Social Forces*.
- FFCWS. 2019. "A Brief Guide to Using the Weights for Waves 1-6." Princeton University; [https://fragilefamilies.princeton.edu/sites/fragilefamilies/files/using\\_the\\_fragile\\_families\\_weights\\_waves\\_1\\_6.pdf](https://fragilefamilies.princeton.edu/sites/fragilefamilies/files/using_the_fragile_families_weights_waves_1_6.pdf).
- Fomby, Paula. 2013. "Family Instability and College Enrollment and Completion." *Population Research and Policy Review* 32 (4): 469–94.
- Fomby, Paula, and Andrew J Cherlin. 2007. "Family Instability and Child Well-Being." *American Sociological Review* 72 (2): 181–204.  
<https://doi.org/10.1177/000312240707200203>.
- Fomby, Paula, and David S Johnson. 2022. "Continuity and Change in US Children's Family Composition, 1968–2017." *Demography*.
- Fomby, Paula, and Stefanie Mollborn. 2017. "Ecological Instability and Children's Classroom Behavior in Kindergarten." *Demography* 54 (5): 1627–51. <https://doi.org/10.1007/s13524-017-0602-2>.
- Fomby, Paula, and Kelly Musick. 2018. "Mothers' Time, the Parenting Package, and Links to Healthy Child Development." *Journal of Marriage and Family* 80 (1): 166–81.
- Fomby, Paula, and Cynthia Osborne. 2017. "Family Instability,

- Multipartner Fertility, and Behavior in Middle Childhood.” *Journal of Marriage and Family* 79 (1): 75–93. <https://doi.org/10.1111/jomf.12349>.
- Fox, Greer Litton, Michael L Benson, Alfred A DeMaris, and Judy Van Wyk. 2002. “Economic Distress and Intimate Violence: Testing Family Stress and Resources Theories.” *Journal of Marriage and Family* 64 (3): 793–807.
- Frazier, E Franklin. 1928. “The Negro Family.” *The ANNALS of the American Academy of Political and Social Science* 140 (1): 44–51.
- Friedman, Jerome H. 2002. “Stochastic Gradient Boosting.” *Computational Statistics & Data Analysis* 38 (4): 367–78.
- Furstenberg, Frank. 2016. “The Changing American Family: An Overview from 1965 to 2015.” In *Changing Family Dynamics and Demographic Evolution: The Family Kaleidoscope*, edited by Dimitri Mortelmans, Koenraad Matthijs, Elisabeth Alofs, and Barbara Segaert, 13–24. Edward Elgar Publishing.
- Garrett, Eilidh, Alice Reid, Kevin Schürer, and Simon Szreter. 2001. *Changing Family Size in England and Wales: Place, Class and Demography, 1891–1911*. Vol. 36. Cambridge University Press.
- Gauthier, Jacques-Antoine, Eric D Widmer, Philipp Bucher, and Cédric Notredame. 2010. “1. Multichannel Sequence Analysis Applied to Social Science Data.” *Sociological Methodology* 40 (1): 1–38. <https://doi.org/10.1111/j.1467-9531.2010.01227.x>.
- Gillies, Val. 2005. “Raising the ‘Meritocracy’ Parenting and the Individualization of Social Class.” *Sociology* 39 (5): 835–53.
- Ginther, Donna K, and Robert A Pollak. 2004. “Family Structure and Children’s Educational Outcomes: Blended Families, Stylized Facts, and Descriptive Regressions.” *Demography* 41 (4): 671–96. <https://doi.org/10.1353/dem.2004.0031>.
- Goisis, Alice, Berkay Özcan, and Philippe Van Kerm. 2019. “Do Children Carry the Weight of Divorce?” *Demography* 56 (3): 785–811.
- Goldfeld, Sharon, Meredith O’Connor, Dan Cloney, Sarah Gray, Gerry Redmond, Hannah Badland, Katrina Williams, et al. 2018. “Understanding Child Disadvantage from a Social Determinants Perspective.” *J Epidemiol Community Health* 72 (3): 223–29.
- Goldthorpe, John H. 2010. “Class Analysis and the Reorientation of Class Theory: The Case of Persisting Differentials in Educational Attainment 1.” *The British Journal of Sociology* 61: 311–35.
- . 2021. “Class and Status in Interwar England: Current Issues in the Light of a Historical Case.” *The British Journal of Sociology* 72 (2): 239–51.
- Goodman, Robert, Tamsin Ford, Helen Simmons, Rebecca Gatward, and Howart Meltzer. 2000. “Using the Strengths and Difficulties Questionnaire (SDQ) to Screen for Child Psychiatric Disorders in a Community Sample.” *The British Journal of Psychiatry* 177 (6): 534–39.

- Gosselin, Julie, Lyzon Babchishin, and Elisa Romano. 2015. "Family Transitions and Children's Well-Being During Adolescence." *Journal of Divorce & Remarriage* 56 (7): 569–89.
- Grimm, H., and H. Doil. 2006. "ELFRA 2-Elternfragebogen für Zweijährige Kinder."
- Grodsky, Eric, John Robert Warren, and Erika Felts. 2008. "Testing and Social Stratification in American Education." *Annu. Rev. Sociol* 34: 385–404. <https://doi.org/10.1146/annurev.soc.34.040507.134711>.
- Gruber, Jonathan. 2004. "Is Making Divorce Easier Bad for Children? The Long-Run Implications of Unilateral Divorce." *Journal of Labor Economics* 22 (4): 799–833.
- Grusec, Joan E. 2011. "Socialization Processes in the Family: Social and Emotional Development." *Annual Review of Psychology* 62: 243–69.
- Guo, Guang, and Kathleen Mullan Harris. 2000. "The Mechanisms Mediating the Effects of Poverty on Children's Intellectual Development." *Demography* 37 (4): 431–47.
- Gupta, Sandeep K. 2011. "Intention-to-Treat Concept: A Review." *Perspectives in Clinical Research* 2 (3): 109.
- Guyon, Herve. 2018. "Variables Latentes Et Propriétés Mentales: Pour Une épistémologie Affirmée Pragmatiste Et réaliste." *Psychologie Française*. <https://doi.org/10.1016/j.psfr.2017.11.002>.
- Hadfield, Kristin, Margaret Amos, Michael Ungar, Julie Gosselin, and Lawrence Ganong. 2018. "Do Changes to Family Structure Affect Child and Family Outcomes? A Systematic Review of the Instability Hypothesis." *Journal of Family Theory & Review* 10 (1): 87–110. <https://doi.org/10.1111/jftr.12243>.
- Hadfield, Kristin, Michael Ungar, and Elizabeth Nixon. 2018. "Rethinking Discourses of Family Instability." *Journal of Family Social Work* 21 (2): 172–87. <https://doi.org/10.1080/10522158.2017.1410268>.
- Hagenaars, Jacques A, and Allan L McCutcheon. 2002. *Applied Latent Class Analysis*. Cambridge University Press.
- Haggerty, Kevin P, Anne McGlynn-Wright, and Tali Klima. 2013. "Promising Parenting Programmes for Reducing Adolescent Problem Behaviours." *Journal of Children's Services*.
- Hall, Matthew, John Iceland, and Youngmin Yi. 2019. "Racial Separation at Home and Work: Segregation in Residential and Workplace Settings." *Population Research and Policy Review*, 1–24. <https://doi.org/10.1007/s11113-019-09510-9>.
- Hällsten, Martin, and Fabian T Pfeffer. 2017. "Grand Advantage: Family Wealth and Grandchildren's Educational Achievement in Sweden." *American Sociological Review* 82 (2): 328–60. <https://doi.org/doi.org/10.1177/0003122417695791>.
- Hansen, Hans-Tore. 2005. "Unemployment and Marital Dissolution: A Panel Data Study of Norway." *European Sociological Review* 21 (2):

135–48.

- Hara, Nobuko. 2016. “Unpaid Labor and the Critique of Political Economy in Home Economics and New Household Economics: From the Feminist Economics Perspective.” *The History of Economic Thought* 58 (1): 1–20.
- Harding, David J, Lisa Gennetian, Christopher Winship, Lisa Sanbonmatsu, and Jeffrey R Kling. 2010. “Unpacking Neighborhood Influences on Education Outcomes: Setting the Stage for Future Research.” National Bureau of Economic Research; <https://www.nber.org/papers/w16055>.
- Härkönen, Juho. 2014. “Emerging Trends in Divorce Research.” *Stockholm Research Reports in Demography 2014* (7), 1–22.
- Härkönen, Juho, Fabrizio Bernardi, and Diederik Boertien. 2017. “Family Dynamics and Child Outcomes: An Overview of Research and Open Questions.” *European Journal of Population* 33 (2): 163–84. <https://doi.org/10.1007/S10680-017-9424-6>.
- Hart, Betty, and Todd R Risley. 1995. *Meaningful Differences in the Everyday Experience of Young American Children*. Paul H Brookes Publishing.
- Harvey, Hope. 2020. “Cumulative Effects of Doubling up in Childhood on Young Adult Outcomes.” *Demography* 57 (2): 501–28.
- Hauser, Robert M, and John Robert Warren. 1997. “Socioeconomic Indexes for Occupations: A Review, Update, and Critique.” *Sociological Methodology* 27 (1): 177–298.
- Haveman, Robert, and Barbara Wolfe. 1995. “The Determinants of Children’s Attainments: A Review of Methods and Findings.” *Journal of Economic Literature* 33 (4): 1829–78.
- Heckman, James J. 2011. “The Economics of Inequality: The Value of Early Childhood Education.” *American Educator* 35 (1): 31.
- Heckman, James J., John Eric Humphries, and Tim Kautz, eds. 2014. *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. Chicago: University of Chicago Press.
- Heckman, James, M Holland, K Makino, D Olds, Rodrigo Pinto, and Maria Rosales-Rueda. 2014. “The Nurse Family Partnership Program: A Reanalysis of the Memphis Randomized Controlled Trial.” University of Chicago Working Paper, ed (June 1, 2014).
- Heene, Moritz. 2013. “Additive Conjoint Measurement and the Resistance Toward Falsifiability in Psychology.” *Quantitative Psychology and Measurement* 4 (246). <https://doi.org/10.3389/fpsyg.2013.00246>.
- Heinrich, Carolyn J. 2014. “Parents’ Employment and Children’s Wellbeing.” *The Future of Children*, 121–46.
- Hernan, M, and J Robbins. 2020. “Causal Inference: What If.” Boca Raton, FL: Chapman & Hall/CRC.
- Hernán, Miguel, Babette Brumback, and James Robins. 2000. “Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men.” *Epidemiology*, 561–70.

- Hernán, Miguel, and James Robins. 2006. "Estimating Causal Effects from Epidemiological Data." *Journal of Epidemiology & Community Health* 60 (7): 578–86. <https://doi.org/10.1136/jech.2004.029496>.
- Hetherington, E Mavis, and Margaret Stanley-Hagan. 1999. "The Adjustment of Children with Divorced Parents: A Risk and Resiliency Perspective." *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 40 (1): 129–40.
- Hill, Jennifer. 2011. "Bayesian Nonparametric Modeling for Causal Inference." *Journal of Computational and Graphical Statistics* 20 (1): 217–40.
- Hill, Jennifer L. 2011. "Bayesian Nonparametric Modeling for Causal Inference." *Journal of Computational and Graphical Statistics* 20 (1): 217–40. <https://doi.org/10.1198/jcgs.2010.08162>.
- Hill, Jennifer, Antonio Linero, and Jared Murray. 2020. "Bayesian Additive Regression Trees: A Review and Look Forward." *Annual Review of Statistics and Its Application* 7: 251–78.
- Hill, Jennifer, and Yu-Sung Su. 2013. "Assessing Lack of Common Support in Causal Inference Using Bayesian Nonparametrics: Implications for Evaluating the Effect of Breastfeeding on Children's Cognitive Outcomes." *The Annals of Applied Statistics*, 1386–1420.
- Hirschman, Daniel. 2016. "Stylized Facts in the Social Sciences." *Sociological Science* 3: 604–26. <https://doi.org/10.15195/v3.a26>.
- Hobbs, Savannah, and Christian King. 2018. "The Unequal Impact of Food Insecurity on Cognitive and Behavioral Outcomes Among 5-Year-Old Urban Children." *Journal of Nutrition Education and Behavior* 50 (7): 687–94.
- Hodapp, Albert F, and Kathryn C Gerken. 1999. "Correlations Between Scores for Peabody Picture Vocabulary Test-III and the Wechsler Intelligence Scale for Children-III." *Psychological Reports* 84 (3\_suppl): 1139–42. <https://doi.org/10.2466/pr0.1999.84.3c.1139>.
- Hoff, Erika. 2003. "The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development via Maternal Speech." *Child Development* 74 (5): 1368–78.
- . 2013. "Interpreting the Early Language Trajectories of Children from Low-SES and Language Minority Homes: Implications for Closing Achievement Gaps." *Developmental Psychology* 49 (1): 4.
- Hoff, Erika, and Brett Laursen. 2019. "Socioeconomic Status and Parenting."
- Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–60.
- Hollingshead, August B et al. 1975. "Four Factor Index of Social Status."
- Humphry, Stephen M. 2013. "A Middle Path Between Abandoning Measurement and Measurement Theory." *Theory & Psychology* 23 (6): 770–85. <https://doi.org/10.1177/0959354313499638>.
- Huque, Md Hamidul, John B Carlin, Julie A Simpson, and Katherine J Lee.

2018. “A Comparison of Multiple Imputation Methods for Missing Data in Longitudinal Studies.” *BMC Medical Research Methodology* 18 (1): 1–16.
- Ishizuka, Patrick. 2019. “Social Class, Gender, and Contemporary Parenting Standards in the United States: Evidence from a National Survey Experiment.” *Social Forces* 98 (1): 31–58.
- Jackson, John W. 2016. “Diagnostics for Confounding of Time-Varying and Other Joint Exposures.” *Epidemiology (Cambridge, Mass.)* 27 (6): 859. <https://doi.org/10.1097/EDE.0000000000000547>.
- Jacoby, Sara F, Laura Tach, Terry Guerra, Douglas J Wiebe, and Therese S Richmond. 2017. “The Health Status and Well-Being of Low-Resource, Housing-Unstable, Single-Parent Families Living in Violent Neighbourhoods in p Hiladelphia, Pennsylvania.” *Health & Social Care in the Community* 25 (2): 578–89. <https://doi.org/10.1111/hsc.12345>.
- Jæger, Mads Meier. 2012. “The Extended Family and Children’s Educational Success.” *American Sociological Review* 77 (6): 903–22. <https://doi.org/10.1177/0003122412464040>.
- Jeynes, William H. 2006. “The Impact of Parental Remarriage on Children: A Meta-Analysis.” *Marriage & Family Review* 40 (4): 75–102. [https://doi.org/10.1300/J002v40n04\\_05](https://doi.org/10.1300/J002v40n04_05).
- Johnson, H. M. 1936. *Pseudo-Mathematics in the Mental and Social Sciences*. Vol. 48. [http://psychology.okstate.edu/faculty/jgrice/psyc3214/Johnson\\_1936\\_PseudoMath.pdf](http://psychology.okstate.edu/faculty/jgrice/psyc3214/Johnson_1936_PseudoMath.pdf); The American Journal of Psychology.
- Kalil, Ariel, and Susan E Mayer. 2016. “Understanding the Importance of Parental Time with Children: Comment on Milkie, Nomaguchi, and Denny (2015).” *Journal of Marriage and Family* 78 (1): 262.
- Kalil, Ariel, Rebecca Ryan, and Elise Chor. 2014. “Time Investments in Children Across Family Structures.” *The ANNALS of the American Academy of Political and Social Science* 654 (1): 150–68.
- Kalil, Ariel, Rebecca Ryan, and Michael Corey. 2012. “Diverging Destinies: Maternal Education and the Developmental Gradient in Time with Children.” *Demography* 49 (4): 1361–83. <https://doi.org/10.1007/s13524-012-0129-5>.
- Kalmijn, Matthijs, and Thomas Leopold. 2020. “A New Look at the Separation Surge in Europe: Contrasting Adult and Child Perspectives.” *American Sociological Review*, 0003122420973982.
- Kang, Joseph DY, Joseph L Schafer, et al. 2007. “Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data.” *Statistical Science* 22 (4): 523–39.
- Karabatsos, George. 2001. “The Rasch Model, Additive Conjoint Measurement, and New Models of Probabilistic Measurement Theory.” *Journal of Applied Measurement* 2 (4): 389–423.

- . 2018. “On Bayesian Testing of Additive Conjoint Measurement Axioms Using Synthetic Likelihood.” *Psychometrika* 83 (2): 321–32. <https://doi.org/10.1007/s11336-017-9581-x>.
- Kautz, Tim, James J Heckman, Ron Diris, Bas Ter Weel, and Lex Borghans. 2014. “Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success.” National Bureau of Economic Research.
- Keller, Heidi, Bettina Lamm, Monika Abels, Relindis Yovsi, Jörn Borke, Henning Jensen, Zaira Papaligoura, et al. 2006. “Cultural Models, Socialization Goals, and Parenting Ethnotheories: A Multicultural Analysis.” *Journal of Cross-Cultural Psychology* 37 (2): 155–72.
- Kelly, Joan B. 2000. “Children’s Adjustment in Conflicted Marriage and Divorce: A Decade Review of Research.” *Journal of the American Academy of Child & Adolescent Psychiatry* 39 (8): 963–73.
- Kelly, Joan B, and Robert E Emery. 2003. “Children’s Adjustment Following Divorce: Risk and Resilience Perspectives.” *Family Relations* 52 (4): 352–62.
- Kennedy, Lauren A, and Andrew Gelman. 2018. “Year 15 Fragile Families Survey Weight Adjustment.”
- Killewald, Alexandra, and Ian Lundberg. 2017. “New Evidence Against a Causal Marriage Wage Premium.” *Demography* 54 (3): 1007–28.
- Kingsbury, Nancy, and John Scanzoni. 2009. “Structural-Functionalism.” In *Sourcebook of Family Theories and Methods*, 195–221. Springer.
- Koot, Hans M, Edwin JCG Van Den Oord, Frank C Verhulst, and Dorret I Boomsma. 1997. “Behavioral and Emotional Problems in Young Preschoolers: Cross-Cultural Testing of the Validity of the Child Behavior Checklist/2-3.” *Journal of Abnormal Child Psychology* 25 (3): 183–96. <https://doi.org/10.1023/A:1025791814893>.
- Kuchirko, Yana. 2019. “On Differences and Deficits: A Critique of the Theoretical and Methodological Underpinnings of the Word Gap.” *Journal of Early Childhood Literacy* 19 (4): 533–62.
- Kuhl, Patricia K. 2004. “Early Language Acquisition: Cracking the Speech Code.” *Nature Reviews Neuroscience* 5 (11): 831–43.
- Kuppens, Sofie, and Eva Ceulemans. 2019. “Parenting Styles: A Closer Look at a Well-Known Concept.” *Journal of Child and Family Studies* 28 (1): 168–81.
- Kuroki, Manabu, and Judea Pearl. 2014. “Measurement Bias and Effect Restoration in Causal Inference.” *Biometrika* 101 (2): 423–37. <https://doi.org/10.1093/biomet/ast066>.
- Kyngdon, Andrew. 2010. “Plausible Measurement Analogies to Some Psychometric Models of Test Performance.” *British Journal of Mathematical and Statistical Psychology* 64 (3): 478–97. <https://doi.org/10.1348/2044-8317.002004>.
- Lacot, Emile, Mohammad H. Afzali, and Stéphane Vautier. 2016. “Test

- Validation Without Measurement.” *European Journal of Psychological Assessment* 32. <https://doi.org/10.1027/1015-5759/a000253>.
- Lahire, Bernard. 2019. *Enfance de Classe*. Seuil.
- Laird, Jessica, Nick Fabrin Nielsen, and Torben Heien Nielsen. 2020. “Differential Effects of the Timing of Divorce on Children’s Outcomes: Evidence from Denmark.”
- Lareau, Annette. 2011. *Unequal Childhoods: Class, Race, and Family Life*. Univ of California Press.
- Lauster, Nathanael T. 2010. “Housing and the Proper Performance of American Motherhood, 1940–2005.” *Housing Studies* 25 (4): 543–57. <https://doi.org/10.1080/02673031003711485>.
- Lazarevic, Vanja, Geneva Toledo, and Jillian Lee Wiggins. 2020. “Influence of Maternal Ethnic–Racial Identity on Children’s Internalizing Symptom Trajectories.” *Journal of Experimental Psychopathology* 11 (1): 2043808719898024.
- Lee, Dohoon, and Sara McLanahan. 2015. “Family Structure Transitions and Child Development: Instability, Selection, and Population Heterogeneity.” *American Sociological Review* 80 (4): 738–63. <https://doi.org/10.1177/0003122415592129>.
- Leśniewska, Justyna, François Pichette, and Sébastien Béland. 2018. “First Language Test Bias? Comparing French-Speaking and Polish-Speaking Participants’ Performance on the Peabody Picture Vocabulary Test.” *Canadian Modern Language Review* 74 (1): 27–52.
- Lesthaeghe, Ron. 1991. *The Second Demographic Transition in Western Countries: An Interpretation*. Interuniversity Programme in Demography.
- . 2014. “The Second Demographic Transition: A Concise Overview of Its Development.” *Proceedings of the National Academy of Sciences* 111 (51): 18112–15.
- Letiecq, Bethany L. 2019. “Surfacing Family Privilege and Supremacy in Family Science: Toward Justice for All.” *Journal of Family Theory & Review* 11 (3): 398–411.
- Lichter, Daniel T, Joseph P Price, and Jeffrey M Swigert. 2019. “Mismatches in the Marriage Market.” *Journal of Marriage and Family*. <https://doi.org/10.1111/jomf.12603>.
- Liddell, Torrin M., and John K. Kruschke. 2018. “Analyzing Ordinal Data with Metric Models: What Could Possibly Go Wrong?” *Journal of Experimental Social Psychology* 79: 328–48. <https://doi.org/10.1016/j.jesp.2018.08.009>.
- Linberg, Anja. 2018. *Interaktion Zwischen Mutter Und Kind: Dimensionen, Bedingungen Und Effekte*. Waxmann Verlag.
- Linberg, Anja, Daniel Mann, Manja Attig, Franziska Vogel, Sabine Weinert, and Hans-Günther Roßbach. 2019. “ASSESSMENT OF INTERACTIONS WITH THE MACRO-ANALYTIC RATING SYSTEM OF PARENT-CHILD-INTERACTIONS IN THE NEPS AT

THE CHILD's AGE OF 7, 17, AND 26 MONTHS."

- Linberg, Tobias, Thorsten Schneider, Jane Waldfogel, and Yi Wang. 2019. "Socioeconomic Status Gaps in Child Cognitive Development in Germany and the United States." *Social Science Research* 79: 1–31.
- Loken, Eric, and Andrew Gelman. 2017. "Measurement Error and the Replication Crisis." *Science* 355 (6325): 584–85. <https://doi.org/10.1126/science.aal3618>.
- Luce, R. Duncan, and John W. Tukey. 1964. "Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement." *Journal of Mathematical Psychology* 1 (1): 1–27. [https://doi.org/10.1016/0022-2496\(64\)90015-X](https://doi.org/10.1016/0022-2496(64)90015-X).
- Ludwig, Volker, and Josef Brüderl. 2018. "Is There a Male Marital Wage Premium? New Evidence from the United States." *American Sociological Review* 83 (4): 744–70.
- Lyngstad, Torkild Hovde. 2011. "Does Community Context Have an Important Impact on Divorce Risk? A Fixed-Effects Study of Twenty Norwegian First-Marriage Cohorts." *European Journal of Population/Revue Européenne de Démographie* 27 (1): 57–77. <https://doi.org/10.1007/s10680-010-9226-6>.
- Maccoby, Eleanor E. 1994. "The Role of Parents in the Socialization of Children: An Historical Overview."
- . 2000. "Parenting and Its Effects on Children: On Reading and Misreading Behavior Genetics." *Annual Review of Psychology* 51 (1): 1–27.
- Macmillan, Ross, and Ronda Copher. 2005. "Families in the Life Course: Interdependency of Roles, Role Configurations, and Pathways." *Journal of Marriage and Family* 67 (4): 858–79.
- Madigan, Sheri, Heather Prime, Susan A Graham, Michelle Rodrigues, Nina Anderson, Jennifer Khoury, and Jennifer M Jenkins. 2019. "Parenting Behavior and Child Language: A Meta-Analysis." *Pediatrics* 144 (4): e20183556.
- Magar, Emily CE, Louise H Phillips, and Judith A Hosie. 2008. "Self-Regulation and Risk-Taking." *Personality and Individual Differences* 45 (2): 153–59.
- Maggi, Stefania, Lori J Irwin, Arjumand Siddiqi, and Clyde Hertzman. 2010. "The Social Determinants of Early Child Development: An Overview." *Journal of Paediatrics and Child Health* 46 (11): 627–35.
- Magis, David, Sébastien Béland, Francis Tuerlinckx, and Paul De Boeck. 2010. "A General Framework and an r Package for the Detection of Dichotomous Differential Item Functioning." *Behavior Research Methods* 42 (3): 847–62. <https://doi.org/10.3758/BRM.42.3.847>.
- Maldonado, Laurie C, and Rense Nieuwenhuis. 2015. "Family Policies and Single Parent Poverty in 18 OECD Countries, 1978–2008." *Community, Work & Family* 18 (4): 395–415.

- <https://doi.org/10.1080/13668803.2015.1080661>.
- Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies* 60 (3): 531–42. <https://doi.org/10.2307/2298123>.
- Mansournia, Mohammad Ali, Mahyar Etminan, Goodarz Danaei, Jay S Kaufman, and Gary Collins. 2017. "Handling Time Varying Confounding in Observational Research." *Bmj* 359: j4587. <https://doi.org/10.1136/bmj.j4587>.
- Manstead, Antony SR. 2018. "The Psychology of Social Class: How Socioeconomic Status Impacts Thought, Feelings, and Behaviour." *British Journal of Social Psychology* 57 (2): 267–91.
- Mari, Luca, Andrew Maul, David Torres Irribarra, and Mark Wilson. 2017. "Quantities, Quantification, and the Necessary and Sufficient Conditions for Measurement." *Measurement* 100: 115–21. <https://doi.org/10.1016/j.measurement.2016.12.050>.
- Masarik, April S, and Rand D Conger. 2017. "Stress and Child Development: A Review of the Family Stress Model." *Current Opinion in Psychology* 13: 85–90. <https://doi.org/10.1016/j.copsyc.2016.05.008>.
- Mathiesen, Kristin S, and Kristian Tambs. 1999. "The EAS Temperament Questionnaire-Factor Structure, Age Trends, Reliability, and Stability in a Norwegian Sample." *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 40 (3): 431–39. <https://doi.org/10.1111/1469-7610.00460>.
- Maul, Andrew. 2017. "Rethinking Traditional Methods of Survey Validation." *Measurement: Interdisciplinary Research and Perspectives* 15 (2): 51–69. <https://doi.org/10.1080/15366367.2017.1348108>.
- Maul, Andrew, David Torres Irribarra, and Mark Wilson. 2016. "On the Philosophical Foundations of Psychological Measurement." *Measurement* 79: 311–20. <https://doi.org/10.1016/j.measurement.2015.11.001>.
- Mayer, Susan E, Ariel Kalil, Philip Oreopoulos, and Sebastian Gallegos. 2019. "Using Behavioral Insights to Increase Parental Engagement the Parents and Children Together Intervention." *Journal of Human Resources* 54 (4): 900–925. <https://doi.org/10.3368/jhr.54.4.0617.8835R>.
- McLanahan, Sara. 2004. "Diverging Destinies: How Children Are Faring Under the Second Demographic Transition." *Demography* 41 (4): 607–27. <https://doi.org/10.1353/dem.2004.0033>.
- McLanahan, Sara, and Wade Jacobsen. 2015. "Diverging Destinies Revisited." In *Families in an Era of Increasing Inequality*, 3–23. Springer.
- McLanahan, Sara, and Christine Percheski. 2008. "Family Structure and the Reproduction of Inequalities." *Annu. Rev. Sociol* 34: 257–76.
- McLanahan, Sara, Laura Tach, and Daniel Schneider. 2013. "The Causal Effects of Father Absence." *Annual Review of Sociology* 39: 399–427. <https://doi.org/10.1146/annurev-soc-071312-145704>.
- McWayne, Christine, Jason T Downer, Rodrigo Campos, and Robby D

- Harris. 2013. "Father Involvement During Early Childhood and Its Association with Children's Early Learning: A Meta-Analysis." *Early Education & Development* 24 (6): 898–922.  
<https://doi.org/10.1080/10409289.2013.746932>.
- Melhuish, Edward C, Mai B Phan, Kathy Sylva, Pam Sammons, Iram Siraj-Blatchford, and Brenda Taggart. 2008. "Effects of the Home Learning Environment and Preschool Center Experience Upon Literacy and Numeracy Development in Early Primary School." *Journal of Social Issues* 64 (1): 95–114.
- Merz, Emily C, Elaine A Maskus, Samantha A Melvin, Xiaofu He, and Kimberly G Noble. 2019. "Socioeconomic Disparities in Language Input Are Associated with Children's Language-Related Brain Structure and Reading Skills." *Child Development*.
- . 2020. "Socioeconomic Disparities in Language Input Are Associated with Children's Language-Related Brain Structure and Reading Skills." *Child Development* 91 (3): 846–60.
- Michell, Joel. 2008. "Is Psychometrics Pathological Science?" *Measurement* 6 (1-2): 7–24. <https://doi.org/10.1080/15366360802035489>.
- Middleton, Joel A, Marc A Scott, Ronli Diakow, and Jennifer L Hill. 2016. "Bias Amplification and Bias Unmasking." *Political Analysis* 24 (3): 307–23.
- Mikolai, Júlia, and Hill Kulu. 2018. "Divorce, Separation, and Housing Changes: A Multiprocess Analysis of Longitudinal Data from England and Wales." *Demography* 55 (1): 83–106.  
<https://doi.org/10.1007/s13524-017-0640-9>.
- Milkie, Melissa A, Kei M Nomaguchi, and Kathleen E Denny. 2015. "Does the Amount of Time Mothers Spend with Children or Adolescents Matter?" *Journal of Marriage and Family* 77 (2): 355–72.  
<https://doi.org/10.1111/jomf.12170>.
- Milkie, Melissa A, Sara B Raley, and Suzanne M Bianchi. 2009. "Taking on the Second Shift: Time Allocations and Time Pressures of US Parents with Preschoolers." *Social Forces* 88 (2): 487–517.
- Miller, Daniel P, Margaret MC Thomas, Maureen R Waller, Lenna Nepomnyaschy, and Allison D Emory. 2020. "Father Involvement and Socioeconomic Disparities in Child Academic Outcomes." *Journal of Marriage and Family* 82 (2): 515–33.
- Millet, Mathias, and Jean-Claude Croizet. 2016. *L'école Des Incapables? La Maternelle, Un Apprentissage de La Domination*. Paris: La Dispute.
- Minh, Anita, Nazeem Muhajarine, Magdalena Janus, Marni Brownell, and Martin Guhn. 2017. "A Review of Neighborhood Effects and Early Child Development: How, Where, and for Whom, Do Neighborhoods Matter?" *Health & Place* 46: 155–74.
- Mishra, Gita, Dorothea Nitsch, Stephanie Black, Bianca De Stavola, Diana Kuh, and Rebecca Hardy. 2009. "A Structured Approach to Modelling

- the Effects of Binary Exposure Variables over the Life Course.” *International Journal of Epidemiology* 38 (2): 528–37.
- Molina, Mario, and Filiz Garip. 2019. “Machine Learning for Sociology.” *Annual Review of Sociology*.
- Mollborn, Stefanie. 2016. “Young Children’s Developmental Ecologies and Kindergarten Readiness.” *Demography* 53 (6): 1853–82.
- Morrison, Donna Ruane, and Andrew J Cherlin. 1995. “The Divorce Process and Young Children’s Well-Being: A Prospective Analysis.” *Journal of Marriage and the Family*, 800–812.
- Moynihan, Daniel Patrick. 1965. “Employment, Income, and the Ordeal of the Negro Family.” *Daedalus*, 745–70.
- Nauck, Bernhard, and Vivian Lotter. 2015. “Parenting Styles and Perceived Instrumentality of Schooling in Native, Turkish, and Vietnamese Families in Germany.” *Zeitschrift für Erziehungswissenschaft* 18 (4): 845–69.
- Newton, Paul E, and Jo-Anne Baird. 2016. “The Great Validity Debate.” Taylor & Francis. <https://doi.org/10.1080/0969594X.2016.1172871>.
- Nguyen, Trang Quynh, Ian Schmid, and Elizabeth A Stuart. 2021. “Clarifying Causal Mediation Analysis for the Applied Researcher: Defining Effects Based on What We Want to Learn.” *Psychological Methods* 26 (2): 255.
- Nielsen, Mark, Daniel Haun, Joscha Kärtner, and Cristine H Legare. 2017. “The Persistent Sampling Bias in Developmental Psychology: A Call to Action.” *Journal of Experimental Child Psychology* 162: 31–38.
- Nieuwenhuis, Rense, and Laurie C Maldonado. 2018. “The Triple Bind of Single-Parent Families: Resources, Employment and Policies,” 1.
- Noble, Kimberly G., Laura E. Engelhardt, Natalie H. Brito, Luke J. Mack, Elizabeth J. Nail, Jyoti Angal, Rachel Barr, William P. Fifer, and Amy J. Elliott. 2015. “Socioeconomic Disparities in Neurocognitive Development in the First Two Years of Life.” *Developmental Psychology* 57 (5): 535–51. <https://doi.org/10.1002/dev.21303>.
- Nomaguchi, Kei, and Melissa A Milkie. 2020. “Parenthood and Well-Being: A Decade in Review.” *Journal of Marriage and Family* 82 (1): 198–223.
- Nomaguchi, Kei, Melissa A Milkie, and Kathleen E Denny. 2016. “Quantity of Maternal Time and Child and Adolescent Development: Response to Kalil and Mayer (2016) and to Waldfogel (2016).” *Journal of Marriage and Family* 78 (1): 269.
- O’Brien, Robert M. 1985. “The Relationship Between Ordinal Measures and Their Underlying Values: Why All the Disagreement?” *Quality & Quantity* 19 (3): 265–77. <https://doi.org/10.1007%2F00170998?LI=true>.
- Ochs, Elinor, and Tamar Kremer-Sadlik. 2015. “How Postindustrial Families Talk.” *Annual Review of Anthropology* 44: 87–103.
- Oppenheimer, Valerie Kincade. 2013. *Work and the Family: A Study in Social Demography*. Elsevier.

- Orthner, Dennis K, Hinckley Jones-Sanpei, and Sabrina A Williamson. 2003. "Family Strength and Income in Households with Children." *Journal of Family Social Work* 7 (2): 5–23.  
[https://doi.org/10.1300/J039v07n02\\_02](https://doi.org/10.1300/J039v07n02_02).
- Pace, Amy, Rufan Luo, Kathy Hirsh-Pasek, and Roberta Michnick Golinkoff. 2017. "Identifying Pathways Between Socioeconomic Status and Language Development." *Annual Review of Linguistics* 3: 285–308.
- Patton, George C, Craig A Olsson, Vegard Skirbekk, Richard Saffery, Mary E Wlodek, Peter S Azzopardi, Marcin Stonawski, et al. 2018. "Adolescence and the Next Generation." *Nature* 554 (7693): 458–66.
- Pearl, Judea. 2009. *Causality*. Cambridge university press.
- Pei, Fei, Xiafei Wang, Susan Yoon, and Erin Tebben. 2019. "The Influences of Neighborhood Disorder on Early Childhood Externalizing Problems: The Roles of Parental Stress and Child Physical Maltreatment." *Journal of Community Psychology* 47 (5): 1105–17.
- Penfield, Randall D., and Gregory Camilli. 2006. "Differential Item Functioning and Item Bias." In *Handbook of Statistics*, 26:125–67. Elsevier. [https://doi.org/10.1016/S0169-7161\(06\)26005-X](https://doi.org/10.1016/S0169-7161(06)26005-X).
- Perkins, Kristin L. 2017. "Household Complexity and Change Among Children in the United States, 1984 to 2010." *Sociological Science* 4: 701–24. <https://doi.org/10.15195/v4.a29>.
- . 2019. "Changes in Household Composition and Children's Educational Attainment." *Demography* 56 (2): 525–48.
- Perkins, Suzanne C, Eric D Finegood, and James E Swain. 2013. "Poverty and Language Development: Roles of Parenting and Stress." *Innovations in Clinical Neuroscience* 10 (4): 10.
- Petersen, Lara Aylin, and Anna-Lena Gerken. 2018. "NEPS Technical Report for Mathematics: Scaling Results of Starting Cohort 1 or Four-Year Old Children." *NEPS Survey Papers*, no. 45.
- Pilkaukas, Natasha, and Katherine Michelmoro. 2019. "The Effect of the Earned Income Tax Credit on Housing and Living Arrangements." *Demography*, 1–24. <https://doi.org/10.1007/s13524-019-00791-5>.
- Price, Joseph. 2010. "The Effect of Parental Time Investments: Evidence from Natural Within-Family Variation." *Unpublished Manuscript*, Brigham Young University.
- Pungello, Elizabeth P, Iheoma U Iruka, Aryn M Dotterer, Roger Mills-Koonce, and J Steven Reznick. 2009. "The Effects of Socioeconomic Status, Race, and Parenting on Language Development in Early Childhood." *Developmental Psychology* 45 (2): 544.
- Quadagno, Jill S et al. 1994. *The Color of Welfare: How Racism Undermined the War on Poverty*. Oxford University Press.
- Raley, R Kelly, and Megan M Sweeney. 2020. "Divorce, Repartnering, and Stepfamilies: A Decade in Review." *Journal of Marriage and Family* 82 (1): 81–99.

- Raley, R Kelly, Inbar Weiss, Robert Reynolds, and Shannon E Cavanagh. 2019. "Estimating Children's Household Instability Between Birth and Age 18 Using Longitudinal Household Roster Data." *Demography*, 1–17. <https://doi.org/10.1007/s13524-019-00806-1>.
- Reichman, Nancy E, Julien O Teitler, Irwin Garfinkel, and Sara S McLanahan. 2001. "Fragile Families: Sample and Design." *Children and Youth Services Review* 23 (4-5): 303–26. [https://doi.org/10.1016/S0190-7409\(01\)00141-4](https://doi.org/10.1016/S0190-7409(01)00141-4).
- Revelle, William, and Richard E. Zinbarg. 2009. "Coefficients Alpha, Beta, Omega, and the GLB: Comments on Sijsma." *Psychometrika* 74 (1): 145–54. <https://doi.org/10.1007/S11336-008-9102-Z>.
- Ridgeway, Greg, Stephanie Ann Kovalchik, Beth Ann Griffin, and Mohammed U Kabeto. 2015. "Propensity Score Analysis with Survey Weighted Data." *Journal of Causal Inference* 3 (2): 237–49.
- Robins, James. 1986. "A New Approach to Causal Inference in Mortality Studies with a Sustained Exposure Period—Application to Control of the Healthy Worker Survivor Effect." *Mathematical Modelling* 7 (9-12): 1393–1512.
- Rodriguez S., Alejandra. 2021. "Fair Comparisons: Life Course Selection Bias and the Effect of Father Absence on US Children." *Advances in Life Course Research*, 100460.
- Romeo, Rachel R, Julia A Leonard, Sydney T Robinson, Martin R West, Allyson P Mackey, Meredith L Rowe, and John DE Gabrieli. 2018. "Beyond the 30-Million-Word Gap: Children's Conversational Exposure Is Associated with Language-Related Brain Function." *Psychological Science* 29 (5): 700–710.
- Rosenbaum, Paul R. 1984. "Testing the Conditional Independence and Monotonicity Assumptions of Item Response Theory." *Psychometrika* 49 (3): 425–35. <https://doi.org/10.1007/BF02306030>.
- Roßbach, Hans-Günther, W. Tietze, and Sabine Weinert. 2005. *Peabody Picture Vocabulary Test-Revised. Deutsche Forschungsversion Des Tests von LM Dunn & LM Dunn von 1981*. Bamberg/Berlin: Pearson.
- Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688.
- Ruggles, Steven. 2015. "Patriarchy, Power, and Pay: The Transformation of American Families, 1800–2015." *Demography* 52 (6): 1797–1823.
- Ryan, Rebecca M, and Amy Claessens. 2013. "Associations Between Family Structure Changes and Children's Behavior Problems: The Moderating Effects of Timing and Marital Birth." *Developmental Psychology* 49 (7): 1219.
- Sachse, S., and W. Von Suchodoletz. 2007. "Diagnostische Zuverlässigkeit Einer Kurzversion Des Elternfragebogens ELFRA-2 Zur Früherkennung von Sprachentwicklungsverzögerungen." *Klinische Pädiatrie* 219: 76–81.

- <https://doi.org/10.1055/s-2006-942174>.
- Saint-Jacques, Marie-Christine, Élisabeth Godbout, Sylvie Drapeau, Toula Kourgiantakis, and Claudine Parent. 2017. "Researching Children's Adjustment in Stepfamilies: How Is It Studied? What Do We Learn?" *Child Indicators Research*, 1–35.  
<https://doi.org/10.1007/s12187-017-9510-7>.
- Savage, Mike, Fiona Devine, Niall Cunningham, Mark Taylor, Li Yaojun, Johs Hjellbrekke, Brigitte. Le Le Roux, Sam Friedman, and Andrew Miles. 2013. "A New Model of Social Class? Findings from the BBC's Great British Class Survey Experiment." *Sociology* 47 (2): 219–50.  
<https://doi.org/10.1177/0038038513481128>.
- Schneider, Daniel, Orestes P Hastings, and Joe LaBriola. 2018. "Income Inequality and Class Divides in Parental Investments." *American Sociological Review* 83 (3): 475–507.
- Schneider, Thorsten, and Tobias Linberg. 2021. "Development of Socio-Economic Gaps in Children's Language Skills in Germany." *Longitudinal and Life Course Studies*.
- Schober, Pia, and Christian Schmitt. 2017. "Day-Care Availability, Maternal Employment and Satisfaction of Parents: Evidence from Cultural and Policy Variations in Germany." *Journal of European Social Policy* 27 (5): 433–46.
- Schultz, Theodore W. 1974. "Fertility and Economic Values." In *Economics of the Family: Marriage, Children, and Human Capital*, 3–22.  
<https://www.nber.org/chapters/c2961.pdf>; University of Chicago Press.
- Schulz, Kenneth F, and David A Grimes. 2005. "Multiplicity in Randomised Trials II: Subgroup and Interim Analyses." *The Lancet* 365 (9471): 1657–61.
- Schweizer, VJ. 2020. "Divorce: More Than a Century of Change, 1900-2018." *Family Profiles, FP-20* 22.
- Seltzer, Judith A. 2019. "Family Change and Changing Family Demography." *Demography* 56 (2): 405–26.  
<https://doi.org/10.1007/s13524-019-00766-6>.
- Sharkey, Patrick, and Jacob W Faber. 2014. "Where, When, Why, and for Whom Do Residential Contexts Matter? Moving Away from the Dichotomous Understanding of Neighborhood Effects." *Annual Review of Sociology* 40: 559–79.  
<https://doi.org/10.1146/annurev-soc-071913-043350>.
- Shear, Benjamin R., and Bruno D. Zumbo. 2014. "What Counts as Evidence: A Review of Validity Studies in Educational and Psychological Measurement." In *Validity and Validation in Social, Behavioral, and Health Sciences*, edited by B. D. Zumbo, Eric K. H., and E. K. H. Chan, 91–111. Springer. [https://doi.org/10.1007/978-3-319-07794-9\\_6](https://doi.org/10.1007/978-3-319-07794-9_6).
- Shi, Baoyi, Christine Choirat, Brent A Coull, Tyler J VanderWeele, and Linda Valeri. 2021. "CMAverse: A Suite of Functions for Reproducible

- Causal Mediation Analyses.” *Epidemiology* 32 (5): e20–22.
- Shonkoff, Jack P. 2011. “Protecting Brains, Not Simply Stimulating Minds.” *Science* 333 (6045): 982–83. <https://doi.org/10.1126/science.1206014>.
- Shonkoff, Jack P, and Dana Charles McCoy. 2021. “Raising the Bar for Evaluating Effectiveness of Early Childhood Interventions.” *Pediatrics* 148 (4).
- Sigle-Rushton, Wendy, and Sara McLanahan. 2004. “Father Absence and Child Well-Being: A Critical Review.” *The Future of the Family* 116: 120–22.
- Sijtsma, Klaas, and L. Andries van der Ark. 2016. “A Tutorial on How to Do a Mokken Scale Analysis on Your Test and Questionnaire Data.” *British Journal of Mathematical and Statistical Psychology* 70 (1): 137–58. <https://doi.org/10.1111/bmsp.12078>.
- Sijtsma, Klaas, and Rob R. Meijer. 2006. “Nonparametric Item Response Theory and Related Topics.” In *Handbook of Statistics*, 26:719–46. [https://doi.org/10.1016/S0169-7161\(06\)26022-X](https://doi.org/10.1016/S0169-7161(06)26022-X).
- Skopek, Jan, and Giampiero Passaretta. 2020. “Socioeconomic Inequality in Children’s Achievement from Infancy to Adolescence: The Case of Germany.” *Social Forces*.
- Smithers, Lisa G, Alyssa CP Sawyer, Catherine R Chittleborough, Neil M Davies, George Davey Smith, and John W Lynch. 2018. “A Systematic Review and Meta-Analysis of Effects of Early Life Non-Cognitive Skills on Academic, Psychosocial, Cognitive and Health Outcomes.” *Nature Human Behaviour* 2 (11): 867–80.
- Smock, Pamela J, and Christine R Schwartz. 2020. “The Demography of Families: A Review of Patterns and Change.” *Journal of Marriage and Family* 82 (1): 9–34.
- Sommer, Anja, and Daniel Mann. 2015. “Qualität Elterlichen Interaktionsverhaltens: Erfassung von Interaktionen Mithilfe Der Eltern-Kind-Interaktions Einschätzskala Im Nationalen Bildungspanel.” *Leibniz-Institut für Bildungsverläufe, Nationales Bildungspanel. Bamberg*.
- Sorkhabi, Nadia, and Jelani Mandara. 2013. “Are the Effects of Baumrind’s Parenting Styles Culturally Specific or Culturally Equivalent?”
- Spengler, Marion, Rodica Ioana Damian, and Brent W Roberts. 2018. “How You Behave in School Predicts Life Success Above and Beyond Family Background, Broad Traits, and Cognitive Ability.” *Journal of Personality and Social Psychology* 114 (4): 620.
- Spera, Christopher. 2005. “A Review of the Relationship Among Parenting Practices, Parenting Styles, and Adolescent School Achievement.” *Educational Psychology Review* 17 (2): 125–46.
- Stockman, Ida J. 2000. “The New Peabody Picture Vocabulary Test-III: An Illusion of Unbiased Assessment?” *Language, Speech, and Hearing Services in Schools* 31 (4): 340–53.
- Straat, J. Hendrik, L. Andries Van der Ark, and Klaas Sijtsma. 2013.

- “Comparing Optimization Algorithms for Item Selection in Mokken Scale Analysis.” *Journal of Classification* 30 (1): 75–99.  
<https://doi.org/10.1007/s00357-013-9122-y>.
- Straus, Murray A. 2017. “Physical Violence in American Families: Incidence Rates, Causes, and Trends.” In *Abused and Battered*, 17–34. Routledge.  
<https://doi.org/10.4324/9781315083131-3>.
- Sullivan, Alice, Sosthenes Ketende, and Heather Joshi. 2013. “Social Class and Inequalities in Early Cognitive Scores.” *Sociology* 47 (6): 1187–1206.
- Tan, Yaoyuan Vincent, and Jason Roy. 2019. “Bayesian Additive Regression Trees and the General BART Model.” *Statistics in Medicine* 38 (25): 5048–69.
- TenEyck, Michael F, Krysta N Knox, and Sarah A El Sayed. 2021. “Absent Father Timing and Its Impact on Adolescent and Adult Criminal Behavior.” *American Journal of Criminal Justice*, 1–25.
- Thaning, Max, and Martin Hällsten. 2020. “The End of Dominance? Evaluating Measures of Socio-Economic Background in Stratification Research.” *European Sociological Review* 36 (4): 533–47.
- Thorpe, Geoffrey L., and Andrej Favia. 2012. “Data Analysis Using Item Response Theory Methodology: An Introduction to Selected Programs and Applications.” *Psychology Faculty Scholarship*, no. 20.
- Trendler, Günter. 2013. “Measurement in Psychology: A Case of Ignoramus Et Ignorabimus? A Rejoinder.” *Theory & Psychology* 23 (5): 591–615.  
<https://doi.org/10.1177/0959354313490451>.
- . 2018. “Conjoint Measurement Undone.” *Theory & Psychology* 29 (1): 100–128. <https://doi.org/10.1177/0959354318788729>.
- Turney, Kristin, and Sarah Halpern-Meekin. 2020. “Parental Relationship Churning and Adolescent Well-Being: Examining Instability Within Families.” *Journal of Marriage and Family*.
- Uher, Jana. 2020. “Measurement in Metrology, Psychology and Social Sciences: Data Generation Traceability and Numerical Traceability as Basic Methodological Principles Applicable Across Sciences.” *Quality & Quantity*, 1–30.
- Uher, Jana, and Elisabetta Visalberghi. 2016. “Observations Versus Assessments of Personality: A Five-Method Multi-Species Study Reveals Numerous Biases in Ratings and Methodological Limitations of Standardised Assessments.” *Journal of Research in Personality* 61: 61–79.
- Ulferts, Hannah. 2020. “Why Parenting Matters for Children in the 21st Century: An Evidence-Based Framework for Understanding Parenting and Its Impact on Child Development.”
- van Buuren, Stef, and Karin Groothuis-Oudshoorn. 2011. “mice: Multivariate Imputation by Chained Equations in r.” *Journal of Statistical Software* 45 (3): 1–67. <https://www.jstatsoft.org/v45/i03/>.
- Van Winkle, Zachary. 2018. “Family Trajectories Across Time and Space:

- Increasing Complexity in Family Life Courses in Europe?" *Demography* 55 (1): 135–64.
- VanderWeele, Tyler. 2015. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press.
- VanderWeele, Tyler J, and Eric J Tchetgen Tchetgen. 2017. "Mediation Analysis with Time Varying Exposures and Mediators." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79 (3): 917–38.
- VanderWeele, Tyler, and Stijn Vansteelandt. 2014. "Mediation Analysis with Multiple Mediators." *Epidemiologic Methods* 2 (1): 95–115.
- VanOrman, Alicia G, and Paola Scommegna. 2016. *Understanding the Dynamics of Family Change in the United States*.  
<https://stage.prb.org/wp-content/uploads/2016/08/prb-population-bulletin-71.1-complex-families-2016.pdf>; Population Reference Bureau.
- Vautier, Stéphane, Michiel Veldhuis, Émile Lacot, and Nadine Matton. 2012. "The Ambiguous Utility of Psychometrics for the Interpretative Foundation of Socially Relevant Avatars." *Theory & Psychology* 22 (6): 810–22. <https://doi.org/10.1177/0959354312450093>.
- Velleman, Paul F., and Leland Wilkinson. 1993. "Nominal, Ordinal, Interval, and Ratio Typologies Are Misleading." *The American Statistician* 47 (1): 65–72. <https://doi.org/10.1080/00031305.1993.10475938>.
- Vernon-Feagans, Lynne, Patricia Garrett-Peters, Michael Willoughby, Roger Mills-Koonce, Family Life Project Key Investigators, et al. 2012. "Chaos, Poverty, and Parenting: Predictors of Early Language Development." *Early Childhood Research Quarterly* 27 (3): 339–51.
- Wagner, Michael. 2020. "On Increasing Divorce Risks." In *Divorce in Europe*, 37–61. Springer, Cham.
- Waldfogel, Jane. 2016. "How Important Is Parental Time? It Depends: Comment on Milkie, Nomaguchi, and Denny (2015)." *Journal of Marriage and Family* 78 (1): 266.
- Wallenborn, Jordyn T, Gregory Chambers, Elizabeth Lowery, and Saba W Masho. 2019. "Marital Status Disruptions and Internalizing Disorders of Children." *Psychiatry Journal* 2019.
- Walper, Sabine. 2021. "Parenting in Germany." Bundesministerium für Familie, Senioren, Frauen und Jugend Referat Öffentlichkeitsarbeit.  
<https://www.bmfsfj.de/resource/blob/183228/4389623f20a8b41f1950a3fa7edce354/parenting-in-germany-9-fb-englisch-data.pdf>.
- Weininger, Elliot B, and Annette Lareau. 2009. "Paradoxical Pathways: An Ethnographic Extension of Kohn's Findings on Class and Childrearing." *Journal of Marriage and Family* 71 (3): 680–95.
- Wells, L Edward, and Joseph H Rankin. 1991. "Families and Delinquency: A Meta-Analysis of the Impact of Broken Homes." *Social Problems* 38 (1): 71–93. <https://doi.org/10.2307/800639>.
- Widom, Cathy Spatz. 2014. "Longterm Consequences of Child

- Maltreatment.” In *Handbook of Child Maltreatment*, 225–47. Springer.
- Wilson, William Julius. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. University of Chicago Press.
- Winkelmann, Rainer, and Stefan Boes. 2006. *Analysis of Microdata*. Zurich: Springer Science & Business Media.
- Womack, Sean R, Melvin N Wilson, Xin Tong, Kathryn Lemery-Chalfant, and Daniel S Shaw. 2022. “Trajectories of Early Childhood Family Instability and the Development of Externalizing Behaviors from Middle Childhood to Adolescence: A Prospective Study of at-Risk Families.” *Child Development*.
- Wood, Robert. 1978. “Fitting the Rasch Model—a Heady Tale.” *British Journal of Mathematical and Statistical Psychology* 31 (1): 27–32. <https://doi.org/10.1111/j.2044-8317.1978.tb00569.x>.
- Wright, Benjamin D. 1992. “Raw Scores Are Not Linear Measures: Rasch Vs. Classical Test Theory CTT Comparison.” *Rasch Measurement Transactions* 6 (1): 208.
- Wu, Lawrence L. 1996. “Effects of Family Instability, Income, and Income Instability on the Risk of a Premarital Birth.” *American Sociological Review*, 386–406.
- Wu, Lawrence L, and Brian C Martinson. 1993. “Family Structure and the Risk of a Premarital Birth.” *American Sociological Review*, 210–32.
- Yao, Liuyi, Zhixuan Chu, Sheng Li, Yaliang Li, Jing Gao, and Aidong Zhang. 2021. “A Survey on Causal Inference.” *ACM Transactions on Knowledge Discovery from Data (TKDD)* 15 (5): 1–46.
- Young, Rebekah, and David R Johnson. 2015. “Handling Missing Values in Longitudinal Panel Data with Multiple Imputation.” *Journal of Marriage and Family* 77 (1): 277–94.
- Zinn, Sabine, Ariane Würbach, Hans Walter Steinhauer, and Angelina Hammon. 2020. “Attrition and Selectivity of the NEPS Starting Cohorts: An Overview of the Past 8 Years.” *AStA Wirtschafts-Und Sozialstatistisches Archiv* 14 (2): 163–206.

## **Abstract**

There are multiple and well-established associations between family behavior - including divorce, repartnering, parenting - and children's wellbeing. However, making a causal interpretation of these associations is not straightforward. In this dissertation, I consider various statistical modeling and measurement issues that complicate the causal attributions made about those associations in the literature in family sociology and social inequality. First, life course informed research suggests that the problem of selection bias in the father absence literature may be more complex than currently thought. After adjusting for multiple time-invariant and -varying confounder covariates, as well as their history, estimates of father absence's effect on children's wellbeing are reduced substantially, a finding which may be refer to as life course selection bias. These results suggest that father absence is mostly a marker of life course cumulative socioeconomic disadvantage, not a cause of negative effects. Second, family instability experienced during childhood, measured by the number of family transitions, is said to negatively affect children's wellbeing. However, time-dependent confounders affected by past episodes of family instability and affecting future family stability might explain away part of the hypothesized negative impact. I show that a dynamic version of the selection hypothesis plays a substantial role countering the family instability hypothesis, and the effects of cumulative family instability are small and not consistent with the family instability hypothesis. Third, research suggest that socioeconomic status gaps in language skills among preschoolers could be substantially reduced by intervening on the parenting styles, practices, and parental investments of low-resource parents. Employing interventional causal mediation analysis, however, and placing attention in the support for some of the assumptions made in the statistical mediation literature, I show parenting mediates around one third of the total effect of SES on early

language skills, but close to nothing of later language skills, which casts doubt on how effective parenting interventions could be. Fourth, the measurement of cognitive abilities, such as language or numeracy skills, is complicated by various features of standardized assessments. Those problems have important implications for the quantification of social inequality in unobservable variables and for causal inference research because test scores capture non-random noise and are therefore biased. The dissertation concludes by making a plea for furthering causal inference thinking in family sociology, social inequality, social mobility, and family demography research. A life-course perspective on the study of the association between family behavior and children's wellbeing reveals that family instability and parenting should be considered as dynamic concepts that vary over time, rather than static ones, thus calling for an appreciation of the true complexity of the family life course.

**Keywords:** life course, causal inference, childhood, adolescence, family

## **Zusammenfassung**

Es gibt zahlreiche und gut belegte Zusammenhänge zwischen familiärem Verhalten - einschließlich Scheidung, Wiederverpartnerung, Erziehungsverhalten - und kindlichem Wohlergehen. Diese Zusammenhänge können jedoch nur begrenzt kausal interpretiert werden. In dieser Dissertation gehe ich auf verschiedene statistische Modellierungs- und Messprobleme ein, die eine kausale Interpretation der in der Literatur zu Familiensoziologie und sozialer Ungleichheit gefundenen Zusammenhängen erschweren. Erstens legt die Lebensverlaufsforchung nahe, dass das Problem der Verzerrung durch Selektion in der Literatur über die Abwesenheit von Vätern komplexer sein könnte als bisher angenommen. Durch die Korrektur von Verzerrungen durch zeitkonstanter und zeitvariabler konfundierender Variablen, sowie vorherigen zeitvariablen konfundierender Variablen, wird die Schätzung des kausalen Effektes der Abwesenheit des Vaters auf das Wohlergehen der Kinder erheblich reduziert. Dieses Ergebnis kann als Lebensverlaufselektionsverzerrung bezeichnet werden. Diese Ergebnisse deuten darauf hin, dass die Abwesenheit des Vaters hauptsächlich ein Indikator für kumulative sozioökonomische Benachteiligungen im Lebensverlauf ist und nicht die Ursache für diese negative Auswirkungen. Zweitens wird in der aktuellen Forschung angenommen, dass familiäre Instabilität in der Kindheit, gemessen an der Zahl der Familienübergänge, das Wohlbefinden der Kinder negativ beeinflusst. Allerdings könnten zeitabhängige konfundierende Faktoren, die durch vergangene Episoden familiärer Instabilität beeinflusst werden und sich auf die künftige Stabilität der Familie auswirken, einen Teil der angenommenen negativen Auswirkungen erklären. Ich zeige, dass eine dynamische Version der Selektionshypothese eine wesentliche Rolle bei der Entkräftung der Hypothese der familiären Instabilität spielt. Die Auswirkungen der kumulativen familiären Instabilität sind gering und

stimmen nicht mit der Hypothese der familiären Instabilität überein, wenn dynamische Selektivität in die Analyse einbezogen wird. Drittens deuten die Forschungsergebnisse darauf hin, dass die soziale Stratifizierung bei den Sprachkenntnissen von Vorschulkindern durch Eingriffe in den Erziehungsstil, die Erziehungspraktiken und die elterlichen Investitionen von Eltern mit wenig Ressourcen erheblich verringert werden könnten. Mit Hilfe einer kausalen Mediationsanalyse und unter Berücksichtigung einiger der in der statistischen Mediationsliteratur getroffenen Annahmen zeige ich jedoch, dass die elterliche Erziehung nur etwa ein Drittel des Gesamteffekts des sozioökonomischen Status auf die frühen Sprachfähigkeiten mediiert, aber fast nichts auf die späteren Sprachfähigkeiten. Dies lässt Zweifel daran aufkommen, wie wirksam eine Intervention in elterliches Erziehungsverhalten sein kann. Viertens wird die Messung kognitiver Fähigkeiten, wie z. B. Sprach- oder Rechenfertigkeiten, durch verschiedene Merkmale standardisierter Beurteilungen erschwert. Diese Probleme haben wichtige Auswirkungen auf die Quantifizierung sozialer Ungleichheit bei unbeobachtbaren Variablen und auf die Forschung zu kausalen Schlussfolgerungen, da Testergebnisse systematisch verzerrt sind. Die Dissertation schließt mit einem Plädoyer zur rigoroseren Anwendung von Methoden der kausalen Inferenz in Familiensoziologie, Familiendemographie und Forschung zu sozialer Ungleichheit und Mobilität.. Eine Lebensverlaufsperspektive bei der Untersuchung des Zusammenhangs zwischen familiärem Verhalten und dem Wohlergehen von Kindern zeigt, dass familiäre Instabilität und elterliche Erziehung als dynamische Konzepte betrachtet werden sollten, die sich im Laufe der Zeit verändern, und nicht als statische Konzepte und daher eine Würdigung der Komplexität des Familienlebenslaufs erfordern.

**Schlüsselwörter:** Lebensverlauf, Kausal Inferenz, Kindheit, Jugend, Familie

### **Selbstständigkeitserklärung**

**Statement of authorship**

Ich erkläre ausdrücklich, dass es sich bei der von mir eingereichten Arbeit um eine von mir selbstständig und ohne fremde Hilfe verfasste Arbeit handelt.  
I expressly declare that the work I have submitted was written independently and without external help.

Ich erkläre ausdrücklich, dass ich sämtliche in der oben genannten Arbeit verwendeten fremden Quellen, auch aus dem Internet (einschließlich Tabellen, Grafiken u. Ä.) als solche kenntlich gemacht habe. Insbesondere bestätige ich, dass ich ausnahmslos sowohl bei wörtlich übernommenen Aussagen bzw. unverändert übernommenen Tabellen, Grafiken o. Ä. (Zitaten) als auch bei in eigenen Worten wiedergegebenen Aussagen bzw. von mir abgewandelten Tabellen, Grafiken o.Ä. anderer Autorinnen und Autoren die Quelle angegeben habe.

I expressly declare that all sources used in the abovementioned work – including those from the Internet (including tables, graphic and suchlike) – have been marked as such. In particular, I declare that I have, without exception, stated the source for any statements quoted verbatim and/or unmodified tables, graphics etc. (i.e. quotations) of other authors.

Mir ist bewusst, dass Verstöße gegen die Grundsätze der Selbstständigkeit als Täuschung betrachtet und entsprechend geahndet werden.

I am aware that violations against the principles of academic independence are considered deception and are punished accordingly.

Datum  
date

Unterschrift Doktorand/in  
signature of doctoral student