
Vliv počtu sourozenců na matematickou, čtenářskou a přírodovědnou gramotnost v ČR*

MARTA HIRSCHOVÁ**

Fakulta filozofická, Západočeská univerzita, Plzeň
Fakulta sociálních studií, Masarykova univerzita, Brno

MARTIN KREIDL**

Fakulta sociálních studií, Masarykova univerzita, Brno

The Effect of Sibship Size on School Achievement in Mathematics, Reading and Science in the Czech Republic

Abstract: This paper investigates the effect of family size on educational achievement of 15-year-old students in the Czech Republic in 2003. The authors find that a continuous measure of the number of co-residing siblings has a negative effect on school achievement and that this effect persists (somewhat reduced in size) even when controlling for family structure, parental education, type of school, and year at school. They find no interactions between family size and other variables in the analysis. The main finding is consistent across various methods of dealing with missing data. The authors conclude that given contemporary population trends, stratification scholars shall in the future pay more attention to model specification and shall include family size among the predictors in status attainment models if they want to obtain unbiased parameter estimates.

Keywords: sibship size, educational achievement, PISA

Sociologický časopis / Czech Sociological Review, 2012, Vol. 48, No. 4: 697–735

1. Úvod

Proměnná velikost rodiny (nebo počet sourozenců) je častou součástí stratifikačních modelů dosahování statusu. Duncan [1968] rozšiřuje tzv. základní model stratifikačního procesu [srov. Blau, Duncan 1967] o proměnnou počet sourozenců již rok po vydání slavné knihy *American Occupational Structure* [Blau, Duncan 1967]. Tato vysvětlující proměnná má podle řady klasických studií nezanedbatel-

* Text vznikl s finanční podporou grantu GA ČR (grant č. P404/11/0130) a výzkumného záměru MŠMT (ev.č. MSM0021622408). Rádi bychom poděkovali recenzentům, neboť v diskusi, kterou recenzní řízení vyvolalo, jsme podstatným způsobem změnili postup analýzy a významně upravili interpretaci získaných dat. Jsme přesvědčeni, že díky těmto změnám vznikl kvalitnější článek.

** Veškerou korespondenci posílejte na adresu: Marta Hirschová, Fakulta filozofická ZČU, Univerzitní 8, 306 14 Plzeň, e-mail: martahirschova@email.cz.

ný vliv na dosažené vzdělání respondenta a zaměstnanecký status v dospělosti [srov. např. Duncan 1968; Featherman, Hauser 1978; Hauser, Sewell 1985; Shavit, Blossfeld 1993]. Podobně byla doložena empirická asociace mezi počtem sourozenců a kognitivním vývojem a mezi počtem sourozenců a školními výsledky dětí [srov. Blake 1981; Ernest, Angst 1983; Hauser, Sewell 1986]. Asociace mezi počtem sourozenců a socioekonomickým postavením je – přinejmenším v západních industrializovaných zemích – brána za přesvědčivě a robustně doloženou. V méně ekonomicky vyspělých zemích nebo ve specifických subpopulacích (definovaných například religiozitou) je tento vztah buď podstatně slabší, nebo není vůbec identifikován, nebo je dokonce kladný [Downey, Neubauer 1998; Gailbraith 1982; Lu, Treiman 2008; Maralani 2006; Shavit, Pierce 1991].¹

Vliv počtu sourozenců (nebo případně dalších aspektů sourozenecké situace) na stratifikační proměnné je v ČR překvapivě málo prozkoumán. V literatuře jsme našli pouze tři publikované empirické studie, které počet sourozenců zahrnují mezi exogenní proměnné. Kreidl [2004, 2008] zohledňuje počet sourozenců v analýze nerovných šancí na vzdělání v několika bývalých socialistických zemích v letech 1948–1989. Používá řadu různých statistických technik (binární logistická regrese, ordinální logistická regrese, multinominální logistická regrese, sekvenční logitový model, podmíněný logitový model) a ukazuje, že efekt počtu sourozenců na vzdělanostní šance byl negativní ve všech zkoumaných kohortách a ve všech zemích bez ohledu na konkrétní použitý statistický model a přesnou specifikaci dalších vysvětlujících proměnných (a případných interakcí). Šafr [2009] kontroluje počet sourozenců i pořadí narození v analýze utváření socioekonomického statusu a distribuce sociálního kapitálu v rodině a ukazuje negativní efekt počtu sourozenců na socioekonomický status (ISEI). V žádné ze zmíněných analýz ale není počet sourozenců hlavním zájmem badatelů.

Cíle tohoto textu jsou následující:

- a. Představit teorie vysvětlující možný vliv počtu sourozenců na kognitivní vývoj a stratifikační postavení.
- b. Doložit empirickou asociaci mezi počtem sourozenců a výsledky standardizovaných testů v současné ČR.
- c. Ověřit, zda zjištěný efekt počtu sourozenců přetrvá i v mnohorozměrném statistickém modelu při kontrole dalších aspektů rodiny původu a studované školy.
- d. Ověřit, zda efekt počtu sourozenců závisí na rodičovské konstelaci a studované škole.

¹ Řada výzkumníků se pokoušela ukázat, že je vzdělání a zaměstnání respondentů ovlivněno i dalšími proměnnými, které charakterizují vztahy mezi sourozenci (např. pořadí narození, věkový rozestup mezi sourozenci a konstelace pohlaví sourozenců [srov. Hauser, Sewell 1985; Hanushek 1992]). Tyto pokusy zaznamenaly v odborné komunitě z řady věcných a metodologických důvodů rezervovanější přijetí [podrobněji Steelman et al. 2002].

2. Proč by mohl počet sourozenců ovlivňovat stratifikační postavení?²

V historii výzkumu efektu počtu sourozenců se postupně objevily tři významné teorie objasňující, proč je v datech možné nalézt asociaci mezi velikostí rodiny a stratifikačním postavením: „confluence theory“ (česky zřejmě nejlépe „soutoková teorie“) [Zajonc, Markus 1975], „resource dilution theory“ (teorie rozdělování zdrojů) [Blake 1986] nebo „alternative theory“ (alternativní teorie) [Guo, VanWey 1999]. První dvě vysvětlení předpokládají kauzální vztah mezi počtem sourozenců a výsledky. Poslední paradigma naopak zpochybňuje kauzalitu tohoto vztahu a zvažuje možnost, že existují charakteristiky (třetí proměnné), které jsou v příčinné souvislosti jak s výsledky dětí, tak s počtem sourozenců (jde tedy o případ nepravého vztahu).

Soutoková teorie je sociálně psychologickým vysvětlením vztahu mezi IQ a velikostí rodiny, pořadím narození a věkovým roztupem. Zajonc a Markus [1975] založili svoji teorii na předpokladu, že s každým dalším narozením dítěte do rodiny se mění interpersonální dynamika rodinného prostředí, což – skrze „učení“ a „učení se“ a kumulaci zkušeností v delším časovém horizontu – ovlivňuje kognitivní vývoj dětí. Tento model pracuje se střední hodnotou úrovně intelektu v rodině, která se vypočítá z průměrných hodnot všech rodinných příslušníků. S každým nově narozeným dítětem se snižuje intelektuální úroveň rodinného prostředí. Protože vývoj dítěte je dán kumulací životních zkušeností, prvorozené dítě je nejdelší dobu vystaveno intelektuálně nejproduktivnějšímu rodinnému prostředí. Zajonc a Markus [1975] z této úvahy dále vyvozují, že jedináčci mají příznivější intelektuální podmínky, a měli by tedy brzy vývojově „předhonit“ děti mající jednoho či více sourozenců.³ „Confluence theory“ byla nejprve sociálními vědci široce přijímána. Později však začaly být diskutovány nedostatky a limity, a dnes je přijímána spíše rezervovaně [např. Hauser, Sewell 1985; Steelman 1985; Blake 1989].

Se vzrůstající kritikou soutokového modelu se více vědců začalo obracet k teorii rozdělování zdrojů, jejímž dominantním vysvětlujícím principem jsou procesy distribuce zdrojů v rodině, což má v důsledku vliv na intelektuální vývoj dětí, vzdělání a další výsledky. Výsledky dětí (školní, intelektuální, stratifikační) závisí na množství zdrojů a na počtu dětí (a současně na věkovém roztupu), mezi které jsou zdroje přerozdělovány. Větší rodina musí více rozdělovat a to zpomaluje intelektuální vývoj dítěte a v konečném důsledku i jeho vzdělanost-

² Označení *stratifikační postavení* chápeme jako zastřešující kategorii pro studijní výsledky, které jsou součástí stratifikačního procesu a představují jeden z determinantů budoucího socioekonomického postavení. Domníváme se, že užití tohoto termínu je přijatelné také proto, že teoretický kontext výzkumu sourozeneckých konstelací se ustavoval v diskuzi vedené obecně v rovině stratifikačního postavení.

³ Na podobném principu Zajonc a Markus [1975] vysvětlují teorii o vyučující funkci, která se týká podobnosti výhod jedináčků a nejstarších dětí a která je založena na myšlence, že starší sourozenec profituje z toho, že učí mladšího sourozence.

ní úspěchy [Blake 1981, 1989]. Tato hypotéza má – ve srovnání se soutokovým modelem – několik zjevných silných stránek: (1) disponuje logickým matematickým nástrojem; (2) představuje široce pojatou teorii, která modifikuje nedostatky soutokového modelu; (3) pokouší se o zařazení dalších teorií jako sociálního kapitálu [Coleman 1988], kulturního kapitálu [Bourdieu 1977] a lidského kapitálu [Becker 1964] a „Quantity-Quality“ modelu [Becker, Tomes 1976] do stratifikačního procesu.

Efekt počtu sourozenců (a tedy i stratifikační a vývojové nevýhody plynoucí z rozdělování omezeného množství zdrojů) je ovšem identifikován pouze tam, kde je zodpovědnost za děti výhradně v rukách nukleární rodiny. Ve společnostech, subkulturách, etnických a náboženských skupinách, v nichž je podstatná míra zodpovědnosti za děti distribuována v širším příbuzenském segmentu, děti z větších rodin znevýhodněny nejsou. Toto platí pro Araby v Izraeli [Shavit, Pierce 1991], pro protestanty v USA [Downey, Neubauer 1998], pro mormony v USA [Gailbraith 1982], pro mladší urbánní a všechny rurální kohorty v Indonésii [Maralani 2008] atp.

Alternativní teorie je reakcí na předchozí výzkumy, které opakovaně zjistily negativní vztah mezi počtem sourozenců a vzdělanostními úspěchy dětí. Model je založen na předpokladu, že kauzální výklad může být sporný v případě, že existují další proměnné, které korelují s počtem sourozenců v rodině a jsou v příčinné souvislosti jak s intelektuálním vývojem dětí, tak s počtem sourozenců.⁴ Guo a VanWey v roce 1999 replikují ve své studii „Sibship Size and Intellectual Development: Is the Relationship Causal?“ výsledky klasické regresní analýzy a používají, na rozdíl od většiny předchozích výzkumů, odhad fixního účinku (*fixed-effects regression*, též tzv. model změny, *change model*).⁵ Aplikují dvě varianty modelů zachycující změny: analýzu sourozenců⁶ a analýzu opakovaných měření [Guo, VanWey 1999].

⁴ Zmínka o tom, že by mohl být negativní vliv počtu sourozenců na úspěch dětí sporný, se objevuje již v dřívějších studiích [např. Ernst, Angst 1983; Steelman 1985; Blake 1989]. Za autory tohoto alternativního výkladu lze však považovat až Gua a VanWeye [Guo, VanWey 1999], kteří jako první zpracovávají tuto problematiku detailně.

⁵ Předpokladem pro analýzu fixního účinku jsou vhodná data, která by splňovala minimálně tato kritéria: (1) intelektuální vývoj dětí je měřen dvakrát v průběhu času na stejných jedincích nebo párech sourozenců; (2) mezi oběma měřeními je dostatečný časový úsek, aby se projevila změna počtu dětí v rodině a (3) měření musí být porovnatelná. V případě Gua a VanWeye výše zmíněné podmínky splňovala data pocházející z národního longitudinálního průzkumu mládeže (NLSY), ve kterých byl intelektuální vývoj dětí měřen v roce 1986 a 1992 [Guo, VanWey 1999].

⁶ Tento model vyžaduje data, která obsahují informace o párech či klastrech sourozenců. Výhoda analýzy sourozenců spočívá především v tom, že je možné kontrolovat efekt genetiky. Pokud je aplikován tradiční přístup (genetický efekt na úrovni sourozenců – dělený mezi vlastní sourozence), je kontrolováno asi 50 % genetického efektu. Výhodnějším řešením je kontrola genetického efektu na úrovni rodičů. Jeden z důvodů, proč preferovat právě genetický efekt na úrovni rodičů, je, že koreluje s velikostí rodiny spíše než genetický

Tato analytická strategie umožňuje lépe než klasická mnohorozměrná regrese kontrolovat neměřené rodinné charakteristiky (socioekonomický status, rodinné intelektuální klima, hodnotový systém rodiny a efekt genetiky), které ovlivňují dovednosti dětí. Jinými slovy tento přístup je schopný kontrolovat všechny konstantní charakteristiky společné pro sourozence v rámci jedné rodiny. Lze tak zároveň identifikovat, kterými charakteristikami se jednotlivé rodiny mezi sebou liší, a nalézt tak společnou příčinu [Guo, VanWey 1999: 172–175]. Guo a VanWey dokládají, že počet sourozenců má malý vliv na verbální dovednosti a dokonce pozitivní vliv na matematické dovednosti. Závěrem shrnují, že po kontrole rodinných charakteristik nemá již počet sourozenců vliv na intelektuální schopnosti dětí [Guo, VanWey 1999: 175–176].

Přestože je tento přístup považován za velmi promyšlený [Phillips 1999: 188] a poměrně dobře zpochybňuje konvenční závěry o vlivu velikosti rodiny na intelektuální vývoj, nevyhnul se ani tento model kritice [Phillips 1999; Downey et al. 1999]. Polemiky poukazují na omezenou interní a externí validitu (matky dětí jsou relativně mladé, ekonomicky znevýhodněné a s nižším vzděláním; analýza je omezena na malé děti; úprava dat před analýzou dále minimalizuje reprezentativitu; model zahrnuje pouze sourozence s velkým věkovým rozestupem) a malou důvěryhodnost některých zjištění (velmi silný pozitivní vliv svobodné matky na čtenářské výsledky dětí, pozitivní vliv počtu sourozenců na výsledky v matematickém testu). Výsledky analýzy se přitom odlišují od jiných analýz, které používaly podobné statistické instrumentarium [Steelman et al. 2002: 255]. Konečně se zdá, že výsledky prezentované Guem a VanWeyem je možné přijatelně interpretovat i s použitím starších, byť modifikovaných, teorií [Phillips 1999].

Hlavním empirickým cílem této studie je ukázat, jaký je v dnešní české společnosti vztah velikosti rodiny (počtu sourozenců) na výsledky patnáctiletých a šestnáctiletých dětí ve standardizovaných testech matematické, čtenářské a přírodovědné gramotnosti. Chceme ukázat, zda tento vztah vůbec existuje, zda přetrvává v mnohorozměrném statistickém modelu i při kontrole dalších charakteristik rodiny původu (rodičovská konstelace, socioekonomický status rodiny) a při kontrole typu školy a ročníku, který žák navštěvuje. Konečně chceme ověřit, zda se efekt počtu sourozenců liší podle rodičovské konstelace (tj. zda existuje statistická interakce mezi počtem sourozenců a rodičovskou situací) a zda případně interaguje s dalšími charakteristikami rodiny žáka.

efekt na úrovni sourozenců. Významnější argument se týká typu rodiny respektive typu sourozenců. Vzhledem k tomu, že vlastní sourozenci mají stejné rodiče (tedy jsou podrobeni stejnému genetickému efektu), je kontrolováno 100 % genetického efektu rodičů místo 50 %. U napůl-vlastních sourozenců je to potom 50 %, protože tito sourozenci sdílejí jen jednoho biologického rodiče [Guo, VanWey 1999: 173–174].

3. Data, proměnné a metody

3.1 Data a způsob zacházení s chybějícími hodnotami

Analýzu provádíme na českých průřezových datech PISA 2003 [UIV 2003],⁷ která jsou součástí mezinárodního výzkumu PISA (*Programme for International Student Assessment*) zaměřujícího se na čtenářskou, matematickou a přírodovědnou gramotnost patnáctiletých a šestnáctiletých žáků. V našem výzkumu pracujeme konkrétně se souborem, který obsahuje data žáků narozených v roce 1987 sesbíraná prostřednictvím mezinárodního žákovského dotazníku rozšířeného o národní proměnné použité v ČR. Soubor navíc zahrnuje data (u stejné skupiny žáků) pocházející z rodičovského dotazníku, který byl sestaven v rámci první vlny longitudinálního výzkumu PISA-L [Sociologický ústav AV ČR 2003].

Longitudinální část dat (rodičovský dotazník) poskytuje nejvíce informací o sourozeneckých konstelacích (počet sourozenců, věk, pořadí narození, pohlaví a genetická spřízněnost sourozenců), které jsou – pokud víme – momentálně v České republice k dispozici. Rodičovský dotazník obsahuje nicméně značný počet chybějících pozorování (1945 dotázaných z celkových 6320 nemá k dispozici vůbec žádné údaje z rodičovského dotazníku; klesá i počet škol, neboť v některých školách žádní žáci vyplněný rodičovský dotazník nepřinesli...). Oproti tomu údaje ze žákovského dotazníku jsou mnohem úplnější (chybějící odpovědi, pokud se u jednotlivých otázek vyskytují, se obvykle týkají jednotek nebo několika málo desítek pozorování), rodinné poměry a sourozenecká situace jsou v něm však měřeny pomocí menšího počtu proměnných s méně podrobnými škálami. Například informace o sourozencích je v žákovském dotazníku získána pomocí otázky „Kdo bydlí u vás doma?“ doprovázené taxativním výčtem možných odpovědí – matka, jiná žena (nevlastní matka nebo pěstounka), otec, jiný muž (nevlastní otec nebo pěstoun), sourozenci, ostatní. Žák kroužkuje tolik možností, kolik pro něj připadá v úvahu, aniž by u jednotlivých osob uváděl další podrobnosti. Omezit analýzu na data ze žákovského dotazníku by tedy znamenalo použít klíčovou vysvětlující proměnnou (počet sourozenců) v dichotomizované podobě: binární proměnná by indikovala, zda respondent má sourozence (žijící ve stejné domácnosti).

Množství chybějících dat v rodičovském dotazníku a hrubé měřicí nástroje v žákovském dotazníku představují klíčový problém pro naši analýzu. Máme pracovat s přesnými měřicími nástroji na omezeném vzorku, nebo – z věcného hlediska – s nepřilíživě přesnými nástroji na úplnějším vzorku? Analytický sociolog pohybující se mezi *Skyllou* nepřilíživě validního měření na přiměřeně reprezentativním vzorku a *Charybdou* validnějšího měření na nereprezentativním vzorku stojí před volbou, kterou nelze řešit podle nějakého standardního a institucionálního postupu. Otázka, zda upřednostnit kritéria externí, nebo interní validity, je v sociologii díky povaze dat zřejmě věčná.

⁷ Data jsou k dispozici na žádost v Ústavu pro informace ve vzdělávání (UIV) u Oddělení mezinárodních výzkumů. Více informací lze získat na webových stránkách www.uiv.cz.

Protože měřicí nástroje byly definovány dlouho před započítím naší analytické práce, jediné „stupně volnosti“, které jsme při zpracování dat měli, se týkaly způsobů zacházení s chybějícími daty. Zdálo se nám přirozené použít kompletní datový soubor ze žákovského dotazníku (i s jeho nedokonalostmi v otázce kvality měření) a použít i neúplná data o počtu sourozenců z rodičovského dotazníku, na která bychom aplikovali různé existující postupy pro práci s chybějícími údaji.

Jaké možnosti zacházení s chybějícími hodnotami jsme zvažovali? V sociologii (nejen české) dlouhou dobu analytici využívali doporučení podpořené například simulačními studiemi Paula Allisona [Allison 2000]: v naprosté většině případů je analýza dostupných pozorování (ve většině softwarových aplikací dostupná pod názvem *listwise deletion*) přijatelnou analytickou strategií, která vede k minimálnímu zkreslení odhadnutých parametrů statistického modelu a jen k mírně zvýšeným standardním chybám. Výjimkou z tohoto pravidla je situace, kdy úplnost vektoru jedné nebo více kovariát (prediktorů) v mnohorozměrném statistickém modelu závisí na hodnotách závisle proměnné. Nejčastější jinou metodou práce s chybějícími daty je dnes v sociálních vědách zřejmě metoda mnohonásobné imputace (doplnění) chybějících hodnot [srov. Allison 2002].

V naší analýze dat z rodičovského dotazníku jsme se rozhodli prezentovat dvě sady výsledků: výsledky založené na analýze všech dat z rodičovského dotazníku (*listwise deletion*) a výsledky založené na mnohonásobně imputovaných datech z rodičovského dotazníku (přičemž imputujeme pouze hodnoty proměnné počet sourozenců). K prezentaci obou sad doplňujících výsledků přistupujeme především proto, abychom se vyhnuli nutnosti upřednostnit *a priori* jeden z výše uvedených možných postupů. Prezentace obou typů analýz nám umožňuje jednak upřednostnit jeden z přístupů až *a posteriori* a zároveň poskytnout prostor čtenáři pro vlastní srovnání a interpretaci. Nezbytnou součástí analýzy všech dostupných pozorování je i analýza reprezentativity dat za úplná pozorování.

3.2 Použité proměnné a metody

V pozici vysvětlované proměnné jsou výsledky testů matematické, čtenářské a přírodovědné gramotnosti. Hlavní vysvětlující proměnnou představuje dichotomická proměnná *přítomnost jednoho a více sourozenců v domácnosti*, případně kontinuální proměnná měřící *počet společně žijících sourozenců*.⁸ Kontrolní pro-

⁸ V analýze používáme pouze proměnné, které referují o sourozencích v domácnosti, a to z toho důvodu, že se jedná o údaj, který poskytuje zároveň žákovský i rodičovský dotazník. Zatímco u rodičovského dotazníku lze rozlišit, zda sourozenci žijí či nežijí v domácnosti (dokonce zde nalezneme i informace o celkovém počtu sourozenců a alokaci vlastních a nevlastních sourozenců), žákovský dotazník toto rozlišení neumožňuje a poskytuje pouze informace o sourozencích v domácnosti. Tento problém představuje poměrně časté omezení řady studií [podrobněji např. Steelman et al. 2002]. Neméně důležitou poznámkou je potom rozsáhlost problematiky rozlišení prostorové alokace sourozenců, jež souvisí

měnné reprezentují skupiny individuálních a kontextuálních charakteristik, které mohou potencionálně ovlivňovat výsledky žáků. Individuální charakteristiky představuje *struktura rodiny* a rodinné zázemí reprezentované *nejvyšším dosaženým vzděláním vzdělanějšího z rodičů*. Kontextuální úroveň reprezentuje *typ školy a navštěvovaný ročník*. Základní soubor zahrnuje 6320 pozorování z 260 škol. Soubor vytvořený pro analýzu dat ze žakovského dotazníku má po úpravě dat a vynechání případů s chybějícími hodnotami kontrolních proměnných celkem 5990 pozorování z 250 škol. Soubor úplných pozorování z rodičovského dotazníku má 4129 pozorování z 243 škol. Soubor imputovaných pozorování z rodičovského dotazníku má 5990 pozorování z 250 škol. Další informace o výzkumu PISA 2003, úpravě datového souboru a popisné statistiky proměnných ze žakovského a rodičovského dotazníku uvádíme v Příloze (tabulka P1 a P2).

Existenci asociace mezi počtem sourozenců a výsledky patnáctiletých a šestnáctiletých žáků v testech gramotnosti zjišťujeme v základní části analýzy pomocí popisných statistik a lineárních regresních modelů (metoda OLS) s robustními standardními chybami upravenými pro shlukovou strukturu dat.⁹ Každou z vysvětlovaných proměnných (matematická, čtenářská a přírodovědná gramotnost) lze považovat za latentní konstrukt popsany pěti hodnotami tzv. *plausible values*, které jsou již součástí původního datového souboru PISA. V analýze používáme obvyklá pravidla pro práci s *plausible values* závisle proměnných, která jsou totožná s tzv. Rubinovými pravidly práce s imputovanými hodnotami [Rubin 1987; srov. též Enders 2010]: hodnota odhadu každého populačního parametru je dána zpřůměrováním dílčích hodnot získaných analýzou každé z *plausible values*.

Celkem v analýze dat ze žakovského dotazníku odhadujeme a prezentujeme 7 modelů pro každou vysvětlovanou proměnnou. První (základní) model vysvětluje výsledky testu pouze existencí sourozenců (Model 1). Prostřednictvím dalších tří mnohorozměrných modelů potom ověřujeme, zda tento vztah přetrvává i po kontrole dalších charakteristik rodiny původu. Kontrolujeme typ rodiny (Model 2), nejvyšší vzdělání vzdělanějšího z rodičů (Model 3) a typ školy a ročník, který žák navštěvuje (Model 4). Mnohorozměrné modely dále rozšiřujeme o interakční efekty. V těchto modelech ověřujeme, zda existuje interakce počtu

s otázkou míry genetické spřízněnosti sourozenců, partnerského uspořádání a struktury rodiny. Jedná se o další rozměr výzkumu sourozeneckých konstelací, a jako takový si zasluhuje samostatné studium. Naše výzkumné zaměření je skromnější, proto tento aspekt ponecháváme stranou s vědomím určitého omezení, které však zásadním způsobem neohrožuje validitu našich výsledků.

⁹ Pozorování (žáky) ze stejné školy nelze v principu považovat za nezávislá [Pfeffermann, Lavange 1989]. Existuje celá řada metod pro zohlednění této skutečnosti v analýze, mimo jiné jde – kromě námi zvoleného postupu – o modely s náhodnými efekty, modely s fixními efekty a marginální modely [srov. Mason 2001]. Námi zvolená metoda vede – při relativně nízkých nárocích na počítačový čas a další zdroje – ke správným závěrům statistického usuzování, tj. výsledky testování hypotéz nejsou shlukovou strukturou dat nežádoucím způsobem ovlivněny.

sourozenců se strukturou rodiny (Model 5), s nejvyšším dosaženým vzděláním vzdělanějšího z rodičů (Model 6) a s typem školy (Model 7).¹⁰

V analýzách úplných dat z rodičovského dotazníku pracujeme se stejnou metodou: OLS regrese s robustními standardními chybami upravenými pro shlukovou strukturu dat. Doplnujeme ji analýzou reprezentativity těchto dat, která je založena na binomiální logistické regresi, v níž závisle proměnná reflektuje, zda daný žák odevzdal vyplněný rodičovský dotazník (a vysvětlujícími proměnnými jsou všechny další znaky používané v předchozích i následujících analýzách). Analýza imputovaných dat z rodičovského dotazníku rovněž pracuje s metodou OLS regrese s robustními standardními chybami korigovanými pro strukturu dat. V analýzách imputovaných dat se soustředíme zejména na to, zda by jiný přístup k definici datového souboru, zacházení s chybějícími daty a měření klíčové vysvětlující proměnné vedl k věcně odlišným závěrům od základní analýzy. Proto v modelech ponecháváme původní binární proměnnou a přidáváme kontinuální vysvětlující proměnnou měřící počet sourozenců. Sledujeme, zda tato dodatečně přidaná proměnná má v modelech významný vliv.

Jednotlivé efekty hodnotíme jak s ohledem na jejich statistickou významnost, tak s ohledem na významnost věcnou. U průměrného skóre považujeme za věcně významné, pokud se liší o více než 2 % od celkového průměru (což je rozdíl přibližně 10 bodů). Za významné rozdíly míry rozptýlenosti potom považujeme ty, které se liší o více než 10 % (přibližně 10 bodů) od celkové směrodatné odchylky. Tento rozdíl považujeme za věcně významný vzhledem k tomu, že v PISE je počítáno bodové skóre metodou IRT (*Item Response Theory*) s použitím Raschovy škály umožňující porovnávání výsledků jednotlivých studentů i skupin studentů. Všechny výše popsané analytické postupy provádíme pomocí statistického softwaru STATA/SE.

4. Empirická zjištění

4.1 Popisné statistiky – data ze žákovského dotazníku

Nejprve předkládáme popisné tabulky evidující efekt jednotlivých proměnných na testové skóre, které doplňujeme o zjištění reprezentující čistý efekt počtu sourozenců v závislosti na dalších individuálních a kontextuálních charakteristikách. Pokud se podíváme na efekt existence sourozenců, tak z hlediska věcného je asociace s výsledky standardizovaných testů nevýznamná (Tabulka 1) – rozdíl průměrného skóre a směrodatné odchylky je minimální. Žáci, kteří mají sourozence, dosahují v průměru pouze o dva body nižšího skóre než jedináčci.¹¹

¹⁰ Rozhodnutí vytvořit modely s interakčními efekty není založené na apriorní teoretické úvaze, ale vychází z dále prezentovaných popisných statistik.

¹¹ Popisné statistiky jsou založené jen na jedné z existujících *plausible values*, konkrétně na páté (v číslování z původního souboru), u každého typu gramotnosti.

Tabulka 1. Průměrné výsledky a směrodatné odchylky (v závorkách) 15 a 16letých žáků v testu matematické, čtenářské a přírodovědné gramotnosti podle počtu sourozenců (žákovský dotazník)

Sourozenci – žákovský dotazník	Gramotnost			N
	matematická	čtenářská	přírodovědná	
Ne	542,25 (93,92)	516,39 (90,04)	550,09 (100,26)	1012
Ano	541,06 (96,18)	512,67 (90,91)	547,59 (101,43)	4978
Celkem	541,26 (95,80)	513,30 (90,77)	548,01 (101,23)	5990

Zdroj: PISA 2003; autoři.

Poznámka: Počet případů N = 5990.

Tabulka 2. Průměrné výsledky a směrodatné odchylky (v závorkách) 15 a 16letých žáků v testu matematické, čtenářské a přírodovědné gramotnosti podle typu rodiny

Typ rodiny	Gramotnost			N
	matematická	čtenářská	přírodovědná	
Biologičtí rodiče	544,28 (96,00)	515,28 (90,79)	550,87 (101,48)	4548
Single rodič	538,86 (90,79)	514,26 (90,60)	550,11 (99,27)	743
Nevlastní rodič	525,38 (98,27)	500,84 (89,52)	528,02 (99,16)	644
Ostatní	510,04 (92,87)	482,50 (91,10)	517,24 (102,62)	55
Celkem	541,26 (95,80)	513,30 (90,77)	548,01 (101,23)	5990

Zdroj: PISA 2003; autoři.

Poznámka: Počet případů N = 5990.

Tabulka 3. Průměrné výsledky a směrodatné odchylky (v závorkách) 15 a 16letých žáků v testu matematické, čtenářské a přírodovědné gramotnosti podle nejvyššího vzdělání vzdělanějšího z rodičů

Nejvyšší vzdělání vzdělanějšího z rodičů	Gramotnost			N
	matematická	čtenářská	přírodovědná	
ZŠ	448,64 (99,23)	431,28 (93,23)	470,07 (106,24)	87
Bez maturity	488,85 (83,76)	467,21 (81,24)	495,6 (90,29)	1271
Maturita	536,1 (86,67)	510,1 (84,20)	541,76 (93,49)	2753
VŠ	588,57 (92,34)	552,97 (87,66)	596,23 (96,55)	1879
Celkem	541,26 (95,80)	513,30 (90,77)	548,01 (101,23)	5990

Zdroj: PISA 2003; autoři.

Poznámka: Počet případů N = 5990.

Naopak jako věcně významný se ukazuje efekt struktury rodiny (Tabulka 2), a to v případě kategorie nevlastní rodič a ostatní. Děti žijící s nevlastním rodičem mají v průměru o 20 bodů (v případě čtenářské gramotnosti 10 bodů) horší výsledky než děti, které bydlí s oběma či jedním biologickým rodičem. V případě ostatních typů rodin se průměrný bodový zisk snižuje o více než 30 bodů, a to u všech typů gramotnosti. Například na škále matematické gramotnosti dosahují děti žijící s oběma biologickými rodiči průměrné hodnoty 544 bodů, děti žijící jen s jedním rodičem průměrné hodnoty 538 bodů, děti žijící s jedním vlastním a jedním nevlastním rodičem průměrné hodnoty 525 bodů a děti žijící v jiném rodinném uspořádání průměrné hodnoty 510 bodů. Na tomto místě je vhodné upozornit na minimální odlišnost směrodatných odchylek u kategorií biologičtí rodiče, single rodič a nevlastní rodič, která svědčí o podobné rozptýlenosti hodnot kolem průměru.

Významné rozdíly mezi průměrným testovým skóre zaznamenáváme také u nejvyššího vzdělání vzdělanějšího z rodičů, kdy roste bodový zisk žáků v průměru o 40 až 50 bodů s každým dalším stupněm vzdělání (Tabulka 3). Například na škále matematické gramotnosti mají děti rodičů se základním vzděláním průměrné skóre 449 bodů, zatímco děti vysokoškolsky vzdělaných rodičů dosahují v průměru 589 bodů. Podobnou markantní diskrepanci nalezneme u čtenářské (431 vs. 553 bodů) a přírodovědné (470 vs. 596 bodů) gramotnosti. Nejnižší míru

Tabulka 4. Průměrné výsledky a směrodatné odchylky (v závorkách) 15 a 16letých žáků v testu matematické, čtenářské a přírodovědné gramotnosti podle typu školy

Typ školy	Gramotnost			N
	matematická	čtenářská	přírodovědná	
ZŠ	499,69 (83,90)	474,32 (82,46)	506,29 (91,08)	2185
Víceleté gymnázium	635,02 (66,58)	599,12 (61,75)	643,46 (71,56)	1158
Čtyřleté gymnázium	612,09 (69,23)	584,79 (63,54)	616,32 (79,76)	398
SOŠ	549,33 (72,59)	521,54 (67,56)	554,56 (80,01)	1529
SOU	460,34 (72,08)	436,58 (67,48)	469,43 (77,38)	720
Celkem	541,26 (95,80)	513,30 (90,77)	548,01 (101,23)	5990

Zdroj: PISA 2003; autoři.

Poznámka: Počet případů N = 5990.

kolísavosti hodnot kolem průměru vykazují děti, kde vzdělanější rodič má nižší než úplné středoškolské vzdělání. Tento rozdíl je věcně významný v porovnání s hodnotami směrodatných odchylek ostatních kategorií.

Rovněž rozdíl mezi efektem jednotlivých typů škol lze považovat z hlediska věcného za významný, a to jak v průměrném skóre, tak ve směrodatných odchylkách (Tabulka 4). Pozoruhodný údaj představuje zejména poměrně vysoká rozptýlenost hodnot kolem průměru u žáků základních škol oproti ostatním typům vzdělávacích institucí. Zejména potom ve srovnání s výrazně nižší kolísavostí hodnot u studentů víceletých gymnázií, mezi kterými jsou zahrnuti žáci kvart (ekvivalent 9. ročníku ZŠ) i kvint (ekvivalent 1. ročníku SOŠ a SOU).

Protože je počet sourozenců nevyhnutelně asociován se strukturou rodiny (rozvedené matky mají v průměru méně dětí než ženy vdané, opakovaně vdané ženy mají v průměru více dětí než ženy v prvním manželství), chceme dále prozkoumat, jaký je vliv počtu sourozenců při kontrole rodinné struktury, což zachycuje Tabulka 5. Vidíme, že počet sourozenců má věcně významný vliv pouze v rodinách s nevlastním rodičem (u matematické gramotnosti je rozdíl v průměrném skóre jedináčku a dětí se sourozenci cca 10 bodů, u čtenářské gra-

Tabulka 5. Čistý efekt počtu sourozenců (žákovský dotazník) po kontrole typu rodiny – průměrné výsledky a směrodatné odchylky (v závorkách) 15 a 16letých žáků v testu matematické, čtenářské a přírodovědné gramotnosti – první část

Sourozenci – žákovský dotazník	Typ rodiny				
	Matematická gramotnost				
	biologičtí rodiče	single rodič	nevlastní rodič	ostatní	celkem
Ne	547,20 (93,76)	539,24 (90,21)	532,72 (100,92)	524,55 (87,42)	542,25 (93,92)
Ano	543,86 (96,33)	538,67 (91,17)	523,05 (97,41)	482,55 (98,97)	541,06 (96,18)
Celkem	544,28 (96,00)	538,86 (90,79)	525,38 (98,27)	510,04 (92,87)	541,26 (95,80)

Poznámka: Počet případů N = 5990.

Tabulka 5. Čistý efekt počtu sourozenců (žákovský dotazník) po kontrole typu rodiny – průměrné výsledky a směrodatné odchylky (v závorkách) 15 a 16letých žáků v testu matematické, čtenářské a přírodovědné gramotnosti – druhá část

Sourozenci – žákovský dotazník	Typ rodiny				
	Čtenářská gramotnost				
	biologičtí rodiče	single rodič	nevlastní rodič	ostatní	celkem
Ne	518,46 (90,60)	516,38 (89,58)	514,33 (88,41)	492,22 (91,29)	516,39 90,04
Ano	514,82 (90,82)	513,22 (91,17)	496,57 (89,54)	464,09 (90,26)	512,67 90,91
Celkem	515,28 (90,77)	514,26 (90,60)	500,84 (89,52)	482,50 (91,10)	513,30 (90,77)

Tabulka 5. Čistý efekt počtu sourozenců (žákovský dotazník) po kontrole typu rodiny – průměrné výsledky a směrodatné odchylky (v závorkách) 15 a 16letých žáků v testu matematické, čtenářské a přírodovědné gramotnosti – dokončení

Sourozenci – žákovský dotazník	Typ rodiny				
	Přírodovědná gramotnost				
	biologičtí rodiče	single rodič	nevlastní rodič	ostatní	celkem
Ne	554,01 (101,97)	551,35 (95,10)	539,35 (99,76)	525,24 (106,31)	550,09 (100,26)
Ano	550,42 (101,41)	549,50 (101,34)	524,43 (98,80)	502,08 (96,16)	547,59 (101,43)
Celkem	550,87 101,48	550,11 99,27	528,02 99,16	517,24 102,62	548,01 (101,23)

Zdroj: PISA 2003; autoři.

Tabulka 6. Čistý efekt počtu sourozenců (žákovský dotazník) po kontrole nejvyššího vzdělání vzdělanějšího z rodičů – průměrné výsledky a směrodatné odchylky (v závorkách) 15 a 16letých žáků v testu matematické, čtenářské a přírodovědné gramotnosti – první část

Sourozenci – žákovský dotazník	Nejvyšší vzdělání vzdělanějšího z rodičů				
	Matematická gramotnost				
	ZŠ	bez maturity	maturita	VŠ	celkem
Ne	501,02 (101,23)	490,15 (81,75)	533,26 (84,66)	593,50 (90,59)	542,25 (93,92)
Ano	434,01 (94,29)	488,60 (84,19)	536,71 (87,11)	587,62 (92,67)	541,06 (96,18)
Celkem	448,64 (99,23)	488,85 (83,76)	536,10 (86,67)	588,57 (92,34)	541,26 (95,80)

Poznámka: Počet případů N = 5990.

Tabulka 6. Čistý efekt počtu sourozenců (žákovský dotazník) po kontrole nejvyššího vzdělání vzdělanějšího z rodičů – průměrné výsledky a směrodatné odchylky (v závorkách) 15 a 16letých žáků v testu matematické, čtenářské a přírodovědné gramotnosti – druhá část

Sourozenci – žákovský dotazník	Nejvyšší vzdělání vzdělanějšího z rodičů				
	Čtenářská gramotnost				
	ZŠ	bez maturity	maturita	VŠ	celkem
Ne	469,73 (92,86)	466,03 (80,09)	510,27 (81,21)	562,25 (87,78)	516,39 (90,04)
Ano	420,54 (91,12)	467,44 (81,48)	510,06 (84,85)	551,17 (87,55)	512,67 (90,91)
Celkem	431,28 (93,23)	467,21 (81,24)	510,10 (84,20)	552,97 (87,66)	513,30 (90,77)

Tabulka 6. Čistý efekt počtu sourozenců (žákovský dotazník) po kontrole nejvyššího vzdělání vzdělanějšího z rodičů – průměrné výsledky a směrodatné odchylky (v závorkách) 15 a 16letých žáků v testu matematické, čtenářské a přírodovědné gramotnosti – dokončení

Sourozenci – žákovský dotazník	Nejvyšší vzdělání vzdělanějšího z rodičů				
	Přírodovědná gramotnost				
	ZŠ	bez maturity	maturita	VŠ	celkem
Ne	525,69 (101,36)	498,24 (86,85)	539,08 (91,77)	603,37 (97,67)	550,09 (100,26)
Ano	454,53 (102,99)	495,11 (90,96)	542,34 (93,86)	594,84 (96,30)	547,59 (101,43)
Celkem	470,07 (106,24)	495,60 (90,29)	541,76 (93,49)	596,23 (96,55)	548,01 (101,23)

Zdroj: PISA 2003; autoři.

motnosti 17 bodů, u přírodovědné gramotnosti 15 bodů) a v ostatních rodinách (rozdíly jsou 42 bodů, 28 bodů a 23 bodů u matematické, čtenářské a přírodovědné gramotnosti). U rodin s oběma biologickými rodiči a u rodin s jedním rodičem je efekt počtu sourozenců minimální (nejvýše 4 body). Tabulka 5 tedy odhaluje interakci mezi strukturou rodiny a počtem sourozenců.

Závislost efektu počtu sourozenců na nejvyšším vzdělání vzdělanějšího z rodičů (Tabulka 6) se projevuje jako věcně významná pouze u dětí, jejichž vzdělanější rodič dosáhl základního nebo vysokoškolského vzdělání. Jedináčci, jejichž vzdělanější rodič dosáhl základního vzdělání, získávají v průměru o 49–71 bodů vyšší skóre u jednotlivých typů gramotnosti než děti, které mají sourozence. Rozptýlenost hodnot kolem průměru je relativně konstantní, a to jak po kontrole typu rodiny, tak po kontrole nejvyššího vzdělání vzdělanějšího z rodičů.¹²

4.2 Mnohorozměrná regrese na datech ze žákovského dotazníku

V Tabulce 7 předkládáme výsledky lineární regresní analýzy. Ve všech modelech, které zde prezentujeme, jsou odhadnuté koeficienty počtu sourozenců statisticky nevýznamné, a to bez ohledu na případné kontrolní proměnné a konkrétní závisle proměnnou. Modely indikují velmi slabý efekt (v intervalu –0,66 až –5,79 podle konkrétního modelu) počtu sourozenců; tyto efekty však v žádném modelu nejsou statisticky významné na hladině 0,05. Celkem tedy i mnohorozměrný statistický model potvrzuje, že v současné české společnosti existence sourozenců nemá čistý aditivní vliv na matematickou, čtenářskou a přírodovědnou gramotnost. Toto zjištění je vzhledem k existující stratifikační literatuře a empirickým výsledkům z celé řady zemí překvapivé.

Výrazné zhoršení testových výsledků můžeme (podle Modelu 2) očekávat u žáků, kteří žijí s nevlastním rodičem a v ostatním typu rodinného uspořádání. Nejvýraznější negativní efekt společného bydlení s nevlastním rodičem se projevuje u testu z matematické a přírodovědné gramotnosti, kdy je výsledné skóre nižší o více než 20 bodů. Po kontrole nejvyššího vzdělání vzdělanějšího z rodičů (Model 3) se koeficienty rodinné struktury u všech gramotností snižují přibližně o 5 až 10 bodů, ale stále ještě zůstávají statisticky významné. Odhadnuté koeficienty u typu rodiny s nevlastním rodičem i u ostatního typu rodin přestávají být statisticky významné po kontrole typu školy a ročníku (Model 4), a to u všech typů gramotnosti. Zde tedy zřejmě jde o nepřímý vliv struktury rodiny, který působí skrze volbu typu školy.

Odhadnuté koeficienty kategorií nejvyššího vzdělání vzdělanějšího z rodičů jsou statisticky významné a tato významnost přetrvává i po kontrole typu

¹² Prověřovali jsme i vztah mezi existencí sourozenců a typem studované školy. Ten zde neprezentujeme v podobě tabulky, pouze konstatujeme, že výsledky existenci této interakce nepotvrzují.

školy a ročníku, a to u matematické a čtenářské gramotnosti. Naopak u přírodovědné gramotnosti přestává být kategorie bez maturity statisticky významná [srov. Straková 2010].

V Tabulce 8 prezentujeme modely s interakčními členy (Model 5, Model 6 a Model 7), které představují rozšířené varianty Modelů 2, 3 a 4. Ze srovnání modelů s interakcemi a bez interakcí vyplývá, že pouze jedna z interakcí je významná – je to interakce mezi existencí sourozenců a vzděláním rodičů (interakce je součástí Modelu 6). Povahu této interakce můžeme posoudit na základě odhadnutých efektů Modelu 6, které jsou prezentovány v Tabulce 8. Vidíme, že negativní efekt počtu sourozenců je velmi silný v rodinách, kde mají rodiče pouze základní vzdělání. V rodinách lépe vzdělaných rodičů je efekt existence sourozenců nevýznamný.

4.3 Analýza úplných dat z rodičovského dotazníku

Tabulka 9 zachycuje odhadnuté efekty počtu sourozenců a dalších proměnných na matematickou, čtenářskou a přírodovědnou gramotnost v modelech OLS regrese odhadnutých na úplných datech z rodičovského dotazníku. Model 8 z této tabulky (ve třech verzích, jedné pro každou závisle proměnnou) je přímou replikací Modelu 4 z Tabulky 7, tj. modelu, který byl odhadnutý na úplných datech z žakovského dotazníku. Vidíme, že ani v jednom z Modelů 8 není efekt binární proměnné „existence sourozenců“ významný (statisticky, ani věcně). Další efekty jsou v Modelech 4 a 8 velmi podobné, s výjimkou odhadnutého efektu vzdělání rodičů, který se na datech z rodičovského dotazníku jeví být podstatně silnějším. V zásadě se zdá, že kdybychom na datech z rodičovského dotazníku aplikovali totožné analytické postupy jako na datech ze žakovského dotazníku, věcný závěr by byl totožný – společné bydlení s jedním nebo více sourozenci se nejeví mít negativní vliv na výsledky standardizovaných testů.

Jak se změní odhadnuté parametry modelů, pokud v úplných datech z rodičovského dotazníku použijeme kontinuální měřítko počtu společně žijících sourozenců? Modely 9 a 10 z Tabulky 9 ukazují tyto situace. Modely 9 používají počet sourozenců jako kontinuální proměnnou, Modely 10 spolu s ní používají i binární indikátor výskytu sourozenců (Model 10 tak testuje linearitu efektu počtu sourozenců). Ve všech třech verzích Modelu 9 je efekt počtu sourozenců negativní (efekty jsou $-5,11$, $-4,96$ a $-4,82$ pro matematickou, čtenářskou a přírodovědnou gramotnost) a statisticky významný na obvyklé pětiprocentní hladině významnosti (příslušné hodnoty hladiny významnosti jsou $0,004$, $0,002$ a $0,020$). Rozdíly mezi Modely 8 a 9 jsou tedy zřetelné a indikují, že použití odlišného měřicího nástroje pro počet sourozenců může mít významné důsledky pro věcné závěry. Modely 10 pak naznačují, proč binární indikátor výskytu sourozenců nepostihuje dostatečně vztah mezi testovými skóry a počtem sourozenců: pokud v modelu současně použijeme binární i intervalové měřítko počtu sourozenců,

Tabulka 7. Odhadnuté koeficienty a robustní standardní chyby (v závorkách) modelů OLS regrese pro výsledky testu matematické, čtenářské a přírodovědné gramotnosti 15 a 16letých žáků – první část

	Matematická gramotnost				Čtenářská gramotnost				Přírodovědná gramotnost			
	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4
Sourozenci – žákovský dotazník (Ne – srov.)												
Ano	-1,83 (3,77)	-4,96 (3,77)	-4,44 (3,44)	-3,31 (2,87)	-3,68 (3,45)	-5,79 (3,53)	-5,27 (3,28)	-4,25 (2,80)	-0,66 (3,99)	-3,34 (3,94)	-2,85 (3,61)	-1,72 (3,13)
Typ rodiny (biologičtí rodiče – referenční sk.)												
Single rodič		-7,12 (4,25)	-4,88 (3,93)	-0,63 (3,35)		-1,11 (3,95)	0,92 (3,66)	4,91 (2,98)		-1,53 (4,20)	0,54 (3,78)	4,83 (3,10)
Nevlastní rodič		-20,49 (4,58)	-14,47 (4,46)	-5,53 (3,52)		-13,29 (4,90)	-8,32 (4,81)	0,38 (4,18)		-22,42 (4,93)	-16,46 (4,80)	-7,57 (4,08)
Ostatní		-32,52 (12,18)	-22,06 (11,71)	-7,27 (10,39)		-38,42 (11,21)	-29,47 (10,41)	-15,11 (9,19)		-38,71 (13,45)	-28,62 (12,36)	-13,91 (10,86)
Nejvyšší vzdělání vzdělanějšího z rodičů (ŽŠ – referenční sk.)												
Bez maturity		35,75 (14,11)	27,49 (12,16)			33,82 (12,99)	25,74 (11,11)			31,80 (16,11)	23,69 (14,59)	
Maturita		83,16 (14,69)	54,75 (12,23)			75,37 (13,36)	47,69 (11,05)			76,61 (16,33)	48,66 (14,42)	
VŠ		135,44 (15,35)	73,65 (12,41)			116,74 (14,13)	57,12 (11,44)			130,13 (17,08)	69,06 (14,92)	

Tabulka 7. Odhadnuté koeficienty a robustní standardní chyby (v závorkách) modelů OLS regrese pro výsledky testu matematické, čtenářské a přírodovědné gramotnosti 15 a 16letých žáků – dokončení

Typ školy (ZŠ – referenční sk.)	Matematická gramotnost				Čtenářská gramotnost				Přírodovědná gramotnost			
	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4
VG												
				114,88 (6,14)				104,08 (5,56)				115,57 (5,80)
4G				89,07 (9,70)				84,12 (7,83)				89,20 (8,69)
SOŠ				36,05 (7,62)				28,38 (6,71)				37,25 (7,06)
SOU				-44,02 (8,15)				-51,34 (7,92)				-40,42 (7,87)
Ročník (9. ročník – stroj.)												
10. ročník				12,26 (4,49)				18,28 (4,93)				11,33 (4,32)
KONSTANTA	542,27 (5,97)	548,28 (6,09)	458,51 (14,93)	451,36 (12,77)	515,27 (5,32)	518,94 (5,62)	439,19 (13,36)	433,07 (11,40)	547,21 (6,02)	552,40 (6,12)	468,22 (17,04)	460,23 (15,65)

Zdroj: PISA 2003; autoři.

Poznámka: Počet případů N = 5990, počet škol N = 250.

Tabulka 8. Odhadnuté koeficienty a robustní standardní chyby (v závorkách) modelů s interakcemi OLS regrese pro výsledky testu matematické, čtenářské a přírodovědné gramotnosti 15 a 16letých žáků – první část

	Matematická gramotnost			Čtenářská gramotnost			Přírodovědná gramotnost		
	M5	M6	M7	M5	M6	M7	M5	M6	M7
Sourozenci – žákovský dotazník (Ne – srov.)									
Ano	-4,58 (4,81)	-56,73 (28,09)	-2,78 (4,52)	-5,57 (4,49)	-54,25 (25,27)	-6,34 (5,14)	-2,44 (5,49)	-60,36 (28,10)	-2,41 (4,56)
Typ rodiny (biologičtí rodiče – referenční sk.)									
Single rodič	-10,96 (7,13)	-4,95 (3,93)	-0,63 (3,35)	-4,42 (7,44)	0,88 (3,65)	4,92 (2,98)	-1,93 (8,51)	0,49 (3,77)	4,27 (3,10)
Nevlastní rodič	-14,28 (9,90)	-14,38 (4,43)	-5,54 (3,53)	-7,89 (8,88)	-8,25 (4,80)	0,37 (4,19)	-18,39 (9,69)	-16,38 (4,77)	-7,58 (4,09)
Ostatní	-22,91 (14,36)	-22,76 (11,72)	-6,99 (10,47)	-33,54 (15,44)	-30,27 (10,28)	-15,01 (9,22)	-31,33 (18,05)	-29,51 (12,35)	-13,94 (10,90)
Nejvyšší vzdělání vzdělanějšího z rodičů (ŽŠ – referenční sk.)									
Bez maturity	-3,60 (22,15)	27,48 (12,21)	25,85 (11,11)	-7,28 (18,44)	25,85 (11,11)	25,85 (11,11)	-15,26 (22,72)	23,71 (14,60)	23,71 (14,60)
Maturita	37,62 (22,37)	54,74 (12,27)	47,78 (11,05)	33,24 (19,37)	47,78 (11,05)	47,78 (11,05)	27,35 (23,61)	48,67 (14,42)	48,67 (14,42)
VŠ	98,47 (23,37)	73,62 (12,45)	57,20 (11,45)	84,05 (20,37)	57,20 (11,45)	57,20 (11,45)	90,64 (25,09)	69,09 (14,93)	69,09 (14,93)

Tabulka 8. Odhadnuté koeficienty a robustní standardní chyby (v závorkách) modelů s interakcemi OLS regrese pro výsledky testu matematické, čtenářské a přírodovědné gramotnosti 15 a 16letých žáků – druhá část

	Matematická gramotnost			Čtenářská gramotnost			Přírodovědná gramotnost		
	M5	M6	M7	M5	M6	M7	M5	M6	M7
Typ školy (ZŠ – referenční sk.)									
VG			116,08 (9,38)			103,86 (8,39)			115,06 (9,22)
4G			91,46 (14,17)			80,14 (12,51)			86,22 (14,91)
SOŠ			37,18 (10,45)			24,65 (8,63)			36,75 (9,82)
SOU			-46,04 (9,55)			-56,11 (10,07)			-41,52 (9,88)
Ročník (9. ročník – stov.)									
10. ročník			12,29 (4,45)			18,34 (4,94)			11,33 (4,29)

Tabulka 8. Odhadnuté koeficienty a robustní standardní chyby (v závorkách) modelů s interakcemi OLS regrese pro výsledky testu matematické, čtenářské a přírodovědné gramotnosti 15 a 16letých žáků – třetí část

INTERAKCE (sourozenci – žákovský dotazník)	Matematická gramotnost		Čtenářská gramotnost		Přírodovědná gramotnost	
	M5	M6	M5	M6	M5	M6
Sourozenci * typ rodiny						
Sourozenci ano * single rodič	5,86 (8,58)		5,02 (9,15)		0,86 (10,73)	
Sourozenci ano * nevlastní rodič	-8,11 (11,20)		-7,08 (11,43)		-5,17 (11,96)	
Sourozenci ano * ostatní	-27,17 (27,72)		-13,79 (26,30)		-20,01 (30,87)	
Sourozenci * nejvyšší vzdělání vzdělanějšího z rodičů						
Sourozenci ano * bez maturity	50,45 (28,63)			52,29 (24,94)		59,99 (27,19)
Sourozenci ano * maturita	57,93 (28,66)			53,62 (25,72)		62,71 (28,81)
Sourozenci ano * VŠ	47,61 (29,10)			42,27 (25,59)		50,95 (28,84)

Tabulka 8. Odhadnuté koeficienty a robustní standardní chyby (v závorkách) modelů s interakcemi OLS regrese pro výsledky testu matematické, čtenářské a přírodovědné gramotnosti 15 a 16letých žáků – dokončení

INTERAKCE (sourozenci – žákovský dotazník)	Matematická gramotnost			Čtenářská gramotnost			Přírodovědná gramotnost		
	M5	M6	M7	M5	M6	M7	M5	M6	M7
Sourozenci * typ školy									
Sourozenci ano * VG			-1,46 (8,30)			0,25 (7,77)			0,62 (8,27)
Sourozenci ano * 4G			-2,96 (9,37)			4,80 (11,35)			3,64 (12,08)
Sourozenci ano * SOŠ			-1,39 (7,65)			4,38 (7,40)			0,61 (7,83)
Sourozenci ano * SOU			2,39 (7,96)			5,70 (9,46)			1,33 (8,29)
KONSTANTA	547,92 (6,68)	499,41 (22,14)	450,93 (12,79)	518,74 (5,99)	477,51 (18,95)	434,70 (11,88)	551,61 (6,98)	513,20 (24,03)	460,77 (15,70)

Zdroj: PISA 2003; autoři.

Poznámka: Počet případů N = 5990, počet škol N = 250.

Tabulka 9. Odhadnuté koeficienty a robustní standardní chyby (v závorkách) modelů OLS regrese pro výsledky testu matematické, čtenářské a přírodovědné gramotnosti 15 a 16letých žáků

	Matematická gramotnost			Čtenářská gramotnost			Přírodovědná gramotnost		
	M8	M9	M10	M8	M9	M10	M8	M9	M10
Počet sourozenců – rodičovský dotazník (0–8)	-5,11 (1,71)	-6,42 (1,71)	(2,12)	-4,96 (1,59)	-5,23 (1,59)	(1,91)	-4,82 (1,97)	-5,82 (1,97)	(2,34)
Sourozenci – rodičovský dotazník (0 – srov.)									
Počet sourozenců > 0	-2,37 (3,94)		7,00 (4,96)	-6,19 (3,78)		1,43 (4,58)	-3,18 (4,53)		5,30 (5,39)
Typ rodiny (Biologičtí rodiče – referenční sk.)									
Single rodič	-5,24 (4,42)	-6,33 (4,42)	-5,68 (4,42)	-0,25 (4,27)	-0,74 (4,25)	-0,60 (4,25)	-1,95 (4,41)	-2,83 (4,42)	-2,35 (4,40)
Nevlastní rodič	-13,55 (5,53)	-13,33 (5,53)	-12,76 (5,56)	-9,40 (5,62)	-8,88 (5,62)	-8,76 (5,68)	-17,23 (5,80)	-16,95 (5,88)	-16,51 (5,84)
Ostatní	-18,41 (14,38)	-18,67 (14,35)	-18,68 (14,24)	-30,39 (14,70)	-30,61 (14,67)	-30,61 (14,65)	-28,03 (15,16)	-28,26 (15,13)	-28,27 (15,06)
Nejvyšší vzdělání vzdělanějšího z rodičů (ZŠ – referenční sk.)									
Bez maturity	28,52 (14,79)	27,10 (14,69)	26,71 (14,72)	29,14 (13,99)	27,74 (13,78)	27,66 (13,78)	22,54 (15,98)	21,19 (15,85)	20,89 (15,90)
Maturita	74,60 (15,16)	71,94 (15,03)	71,33 (15,07)	68,51 (13,58)	65,98 (13,34)	65,86 (13,34)	64,99 (15,71)	62,49 (15,55)	62,03 (15,60)
VŠ	129,66 (15,73)	126,77 (15,61)	126,10 (15,66)	110,82 (14,63)	108,07 (14,37)	107,93 (14,38)	119,77 (16,74)	117,05 (16,56)	116,55 (16,63)
KONSTANTA	473,43 (15,37)	480,41 (15,01)	476,49 (15,25)	455,19 (13,92)	458,48 (13,37)	457,68 (13,65)	488,97 (15,95)	494,71 (15,89)	491,75 (15,83)

Zdroj: PISA 2003; autoři.

Poznámka: Počet případů N = 4129, počet škol N = 243.

je čistý efekt kontinuální proměnné v Modelu 10 silnější než v Modelu 9 a čistý efekt binární proměnné je kladný (byť jde o efekty statisticky nevýznamné). Modely 10 tedy indikují určitou nelinearitu ve vztahu mezi počtem sourozenců a testovými skóry: výsledky v testech jsou pro jedináčky zřejmě významně nižší, než by odpovídalo lineární predikci (podobně je tomu v řadě dalších zemí [srov. Park 2008]).

4.4 Analýza ne-existence dat z rodičovského dotazníku

Výsledky z Modelů 8, 9 a 10 tedy ukazují, že by zjištění získaná z dat z rodičovského dotazníku mohla být významně odlišná od zjištění z dat žákovského dotazníku, a to zřejmě proto, že vztah mezi počtem sourozenců a testovými skóry není lineární. Jak důvěryhodné ale tyto výsledky jsou ve světle nízké návratnosti rodičovského dotazníku? Je rodičovský dotazník dostatečně reprezentativním, abychom výše uvedené výsledky mohli brát vážně, nebo je chybějících dat příliš mnoho, či je dokonce jejich struktura taková, že může významně zkreslit správnost empirických zjištění?

Odpověď na tyto otázky dávají modely binární logistické regrese prezentované v Tabulce 10. Jde o modely, které predikují, zda žáci odevzdali nebo neodevzdali vyplněný rodičovský dotazník (1 – neodevzdali, 0 – odevzdali). Jako prediktory používáme všechny vysvětlující proměnné z dosavadních modelů (Model 11) doplněné o hodnotu příslušného testového skóru (Modely 12, 13 a 14). Už první z logistických modelů ukazuje, že výskyt chybějících dat nesplňuje ani jeden z „jednoduše“ řešitelných vzorců chybějících dat. Data nechybí na základě zcela náhodného procesu (*missing completely at random* – MCAR¹³), což je situace, kdy lze problém chybějících dat vlastně zcela ignorovat a přesto získat nezkreslené odhady populačních parametrů. Z dat samotných přirozeně nepoznáme, zda jsou chybějící hodnoty *výsledkem náhodného procesu* (*missing at random* – MAR), protože může existovat hypotetická neměřená, nebo do modelu nezahrnutá proměnná, která je asociována s výskytem chybějících hodnot. I kdyby ale žádná taková vynechaná proměnná neexistovala a data skutečně splňovala podmínku MAR, z modelů jednoznačně vyplývá, že by rozhodně šlo o situaci MAR se závislostí na hodnotě závisle proměnné. To je učebnicová situace, kdy nelze při práci s chybějícími daty spoléhat ani na metodu *listwise deletion* [Allison 2000]. Je tedy zřejmé, že závěry z Modelů 8, 9 a 10 tak nelze kvůli povaze chybějících dat přijmout jako spolehlivé.

¹³ Odborná literatura rozlišuje chybějící data ne-ignorovatelná (*non-ignorable*, nebo také *not missing at random*, což je stav, kdy výskyt chybějící hodnoty nějaké proměnné závisí na hodnotě té samé proměnné), *missing completely at random* (kdy je pravděpodobnost výskytu chybějících dat rovna u všech pozorování konstantě) a *missing at random*. Poslední situace MAR je situace tzv. informativních chybějících dat, tj. jde o stav, kdy výskyt chybějícího údaje závisí (pouze) na hodnotách dalších proměnných v datech [viz např. Allison 2002].

Tabulka 10. Odhadnuté koeficienty a standardní chyby (v závorkách) modelů logistické regrese nevyplnění rodičovského dotazníku

	M11	M12	M13	M14
Typ rodiny (biologičtí rodiče – referenční sk.)				
Single rodič	0,10 (0,10)	0,01 (0,10)	0,03 (0,10)	0,02 (0,10)
Nevlastní rodič	0,41 (0,10)	0,39 (0,10)	0,42 (0,10)	0,39 (0,10)
Ostatní	0,46 (0,31)	0,44 (0,32)	0,42 (0,32)	0,44 (0,31)
Nejvyšší vzdělání vzdělanějšího z rodičů (ZŠ – referenční sk.)				
Bez maturity	-0,32 (0,26)	-0,27 (0,26)	-0,27 (0,26)	-0,28 (0,26)
Maturita	-0,50 (0,26)	-0,40 (0,26)	-0,41 (0,26)	-0,41 (0,26)
VŠ	-0,41 (0,26)	-0,27 (0,27)	-0,29 (0,27)	-0,27 (0,26)
Typ školy (ZŠ – referenční sk.)				
VG	-1,04 (0,26)	-0,72 (0,27)	-0,72 (0,27)	-0,76 (0,27)
4G	-1,00 (0,40)	-0,77 (0,40)	-0,76 (0,40)	-0,79 (0,40)
SOŠ	-0,46 (0,27)	-0,38 (0,27)	-0,41 (0,27)	-0,39 (0,27)
SOU	-0,09 (0,31)	-0,22 (0,31)	-0,27 (0,31)	-0,20 (0,31)
Ročník (9. ročník – srov.)				
10. ročník	-0,19 (0,17)	-0,14 (0,17)	-0,11 (0,17)	-0,15 (0,16)
Testové skóre				
Matematika (* 100)		-2,78 (0,48)		
Čtení (* 100)			-3,16 (0,48)	
Přírodní vědy (* 100)				-2,49 (0,42)
KONSTANTA	-0,15 (0,28)	1,14 (0,36)	1,25 (0,35)	1,01 (0,34)

Zdroj: PISA 2003; autoři.

Poznámka: Počet případů N = 5990, počet škol N = 250.

4.5 Analýza mnohonásobně imputovaných dat z rodičovského dotazníku

Jsme tedy v situaci, kdy chybějící data v rodičovském dotazníku mohou být MAR, ale víme, že spolehlivým řešením není analýza úplných dat, protože neúplnost dat závisí, mimo jiné, na testových skóre. Můžeme zároveň důvodně očekávat, že nejde o „ne-ignorovatelná“ (*non-ignorable*) chybějící data, protože je jen těžké si představit, že by nevyplnění rodičovského dotazníku záviselo na počtu sourozenců v domácnosti. Proto je myslitelné počítat dále s předpokladem, že data o počtu společně bydlících sourozenců jsou MAR. To je situace, kdy se lze při odhadech populačních parametrů poměrně dobře spolehnout na metodu mnohonásobné imputace [Allison 2000]. Proto v dalším kroku doplníme chybějící údaje o sourozencích z rodičovského dotazníku a provedeme analýzu vztahu mezi testovými skóre a imputovaným počtem sourozenců.

Proces doplnění chybějících hodnot metodou mnohonásobné imputace zahrnuje několik zásadních analytických rozhodnutí. V naší situaci jde zejména o výběr statistického modelu, kterým budeme chybějící hodnoty dopočítávat, a způsob, kterým do procesu imputace zahrneme v datech se přirozeně vyskytující shluky podobných pozorování (žáky stejné školy). Další rozhodnutí nutná a obvyklá v procesu imputace zde podrobněji nekomentujeme; pouze konstatujeme, že jsme se rozhodli imputovat jen jednu proměnnou z rodičovského dotazníku (počet sourozenců), a to pomocí binární proměnné pro výskyt sourozenců ze žákovského dotazníku, typu rodinné struktury, vzdělání rodičů, typu školy, ročníku, pohlaví a všech existujících *plausible values* pro skóre z matematiky, čtení i přírodních věd, a nakonec jsme zahrnuli i sadu dichotomických indikátorů příslušnosti k jednotlivým školám (viz níže).¹⁴

Data doplňujeme metodou *predictive mean matching*, což je jedna z metod vhodných pro doplňování kontinuálních proměnných. V naší situaci má tato metoda dvě výhody: respektuje diskrétní povahu původní proměnné (počty sourozenců mohou být jen celá čísla) a respektuje původní distribuci hodnot imputované proměnné (doplněné hodnoty tedy nemohou být nižší než 0 a vyšší než 8, což byl maximální počet sourozenců uvedený v rodičovském dotazníku).

Otázka zohlednění shlukové struktury dat v imputačním procesu je složitější a nemá zřejmě jednoznačně přijímané řešení [srov. Eddings, Marchenko 2010]. Rozhodli jsme se použít dichotomizovaný indikátor příslušnosti ke škole jako jednu z proměnných, na jejichž základě jsou imputované hodnoty získávány. Tuto metodu jsme použili navzdory tomu, že je obvykle doporučována jen v situacích,

¹⁴ Neznáme strukturu původních dat před výpočtem *plausible values*, takže neznáme ani původní strukturu chybějících hodnot (pokud se tyto v datech vyskytovaly). Předpokládáme, že latentní (skutečné) hodnoty schopností v matematice, čtení a přírodních vědách jsou chybějící a že byly dopočteny pětinasobnou imputací. Vytváříme celkem pět imputovaných hodnot pro počet sourozenců. Pokud by šlo čistě o naši osobní volbu, preferovali bychom vyšší počet imputací. Protože ale v datech již je 5 *plausible values* pro každé testové skóre, našich pět imputací se snadno s původním datovým souborem spojí.

kdy je v datech jen několik málo desítek přirozeně existujících klusterů (v našich datech je 250 škol = shluků dat).

Pomocí imputace jsme byli schopni pokračovat v práci se všemi 5990 respondenty, na nichž jsme prováděli první část analýzy. Navíc jsme schopni použít dodatečnou proměnnou z rodičovského dotazníku, která měří počet společně bydlících sourozenců na škále od 0 do 8. V analýze prezentujeme postupně tři modely pro každou závisle proměnnou (modely jsou prezentovány v Tabulce 11 a 12). Model 15 používá počet sourozenců jako kontinuální proměnnou společně se vzděláním rodičů a rodinnou strukturou (jde o variantu Modelu 3 odhadnutou na datech z rodičovského dotazníku), Model 16 používá současně binární prediktor indikující, zda má respondent sourozence, i kontinuální proměnnou „počet sourozenců“ (a rovněž kontroluje vzdělání rodičů a rodinnou strukturu) a poslední dva Modely 17 a 18 ukazují, zda i na imputovaných datech z rodičovského dotazníku identifikujeme interakci mezi počtem sourozenců a rodinnou strukturou a počtem sourozenců a vzděláním rodičů. Model 15 slouží pro srovnání, do jaké míry se věcné závěry ohledně efektu počtu sourozenců na testová skóre budou lišit, pokud použijeme odlišné operacionalizace počtu sourozenců. Model 16 pak ukazuje, zda při současném použití binární i kontinuální proměnné přináší kontinuální proměnná nějaké nové věcné zjištění.

Zjištění z Modelů 15 a 16 podporují tezi, že počet sourozenců ovlivňuje výsledky testů (a to i při kontrole vzdělání rodičů a typu rodiny původu). Model 15 ukazuje, že u matematické gramotnosti s každým dodatečným společně bydlícím sourozencem klesá očekávaná hodnota testového skóre o 4,82 bodu. U čtenářské gramotnosti je efekt $-4,94$ a u přírodovědné gramotnosti jde o $-4,78$. Všechny tyto efekty jsou statisticky významné na hladině 0,05 (jednotlivé p-hodnoty jsou 0,002, 0,000 a 0,005).

Model 15 neindikuje žádnou podstatnou změnu v efektech dalších proměnných: průměrné skóre u studentů z nevlastních a jiných rodin jsou významně nižší než testové skóre studentů žijících s oběma vlastními rodiči nebo s jedním vlastním rodičem (v případě nevlastních rodičů je skóre v průměru nižší o 8 až 16 bodů, podle závisle proměnné, u „ostatních“ rodin je rozdíl 20 až 27 bodů). Rovněž efekt vzdělání zůstává velmi silný, např. rozdíl v průměrném testovém skóre z matematiky mezi žáky, jejichž rodiče mají základní vzdělání, a žáky, jejichž rodiče mají vysokoškolské vzdělání, je 133 bodů.

Model 16 potom ukazuje vzájemný vztah mezi kontinuální a binární proměnnou odvozenou z počtu sourozenců. Vidíme, že při společném použití obou proměnných efekt kontinuální proměnné mírně roste. Efekt binární proměnné nicméně ani v jedné verzi Modelu 16 není statisticky ani věcně významný (p-hodnoty jsou 0,669, 0,834 a 0,396, pro matematickou, čtenářskou a přírodovědnou gramotnost). Tato nevýznamná binární proměnná tedy neindikuje žádnou zásadní odchylku od linearity ve vztahu mezi počtem sourozenců a testovými skóry.

Protože analýza dat z žákovského dotazníku indikovala interakci mezi existencí sourozenců a strukturou rodiny a interakci mezi vzděláním rodičů a po-

Tabulka 11. Odhadnuté koeficienty a robustní standardní chyby (v závorkách) modelů OLS regrese pro výsledky testu matematické, čtenářské a přírodovědné gramotnosti 15 a 16letých žáků

	Matematická gramotnost		Čtenářská gramotnost		Přírodovědná gramotnost	
	M15	M16	M15	M16	M15	M16
Počet sourozenců – rodičovský dotazník (0–8)						
	–4,82	–5,10	–4,94	–5,08	–4,78	–5,39
	(1,51)	(1,73)	(1,37)	(1,67)	(1,64)	(1,91)
Sourozenci – žákovský dotazník (Ne – srov.)						
Ano		1,71		0,85		3,64
		(3,98)		(4,05)		(4,26)
Typ rodiny (biologičtí rodiče – referenční sk.)						
Single rodič	–5,36	–5,10	0,56	0,70	–0,26	0,31
	(3,96)	(3,94)	(3,60)	(3,66)	(3,78)	(3,79)
Nevlastní rodič	–14,03	–13,85	–7,79	–7,70	–16,21	–15,80
	(4,45)	(4,44)	(4,68)	(4,79)	(4,79)	(4,75)
Ostatní	–19,99	–19,10	–26,96	–26,51	–27,36	–25,49
	(11,56)	(11,76)	(10,48)	(10,50)	(12,58)	(12,69)
Nejvyšší vzdělání vzdělanějšího z rodičů (ZŠ – referenční sk.)						
Bez maturity	34,47	34,39	32,48	32,42	30,54	30,36
	(14,03)	(14,06)	(13,07)	(13,10)	(16,06)	(16,09)
Maturita	80,78	80,66	72,93	72,85	74,23	73,97
	(14,61)	(14,65)	(13,47)	(13,51)	(16,28)	(16,31)
VŠ	132,82	132,68	114,04	113,95	127,52	127,19
	(15,27)	(15,31)	(14,19)	(14,25)	(16,99)	(17,03)
KONSTANTA	463,03	462,02	443,22	442,73	474,07	471,93
	(14,82)	(14,94)	(13,50)	(13,45)	(16,71)	(17,01)

Zdroj: PISA 2003; autoři.

Poznámka: Počet případů N = 5990, počet škol N = 250.

Tabulka 12. Odhadnuté koeficienty a robustní standardní chyby (v závorkách) modelů OLS regrese pro výsledky testu matematické, čtenářské a přírodovědné gramotnosti 15 a 16letých žáků – první část

	Matematická gramotnost		Čtenářská gramotnost		Přírodovědná gramotnost	
	M17	M18	M17	M18	M17	M18
Počet sourozenců – rodičovský dotazník (0–8)	–9,28 (1,87)	–7,37 (11,08)	–8,83 (1,89)	–10,28 (8,67)	–9,12 (2,12)	–9,65 (9,14)
Typ rodiny (biologičtí rodiče – referenční sk.)						
Single rodič	–10,79 (6,29)		–2,88 (5,43)		–4,46 (5,81)	
Nevlastní rodič	–22,57 (7,15)		–17,28 (5,92)		–26,02 (7,22)	
Ostatní	–26,87 (22,58)		–35,98 (18,96)		–33,92 (20,39)	
Nejvyšší vzdělání vzdělanějšího z rodičů (ZŠ – referenční sk.)						
Bez maturity		32,27 (23,26)		23,73 (21,24)		24,02 (21,67)
Maturita		79,31 (23,95)		65,54 (20,89)		68,49 (21,75)
VŠ		131,10 (24,95)		108,30 (21,87)		123,15 (23,90)

čtem sourozenců, prověřujeme existenci těchto interakcí i na doplněných datech z rodičovského dotazníku, a to konkrétně v Modelech 17 a 18 (existují tři varianty modelu, jedna pro každou ze závisle proměnných), jež prezentujeme v Tabulce 12. Jak je z Tabulky 12 zřejmé, v žádné verzi Modelu 17 není interakce mezi počtem sourozenců a strukturou rodiny významná. Podobně v Modelech 18 není významná interakce mezi vzděláním rodičů a počtem sourozenců. Identifikovaný hlavní efekt počtu sourozenců tedy platí ve všech druzích původních rodin a při všech úrovních rodičovského vzdělání do stejné míry (odhady z Modelů 17 ukazují, že s každým dodatečným sourozencem klesá průměrné testové skóre o přibližně 9 bodů, odhady z Modelů 18 indikují efekt v intervalu od –7,37 po –10,28).

Tabulka 12. Odhadnuté koeficienty a robustní standardní chyby (v závorkách) modelů OLS regrese pro výsledky testu matematické, čtenářské a přírodovědné gramotnosti 15 a 16letých žáků – dokončení

	Matematická gramotnost		Čtenářská gramotnost		Přírodovědná gramotnost	
	M17	M18	M17	M18	M17	M18
Interakce						
Single rodič * počet sourozenců	2,11 (3,99)		-0,49 (4,27)		1,08 (4,12)	
Nevlastní rodič * počet sourozenců	2,08 (4,32)		3,64 (4,28)		3,10 (4,71)	
Ostatní rodiny * počet sourozenců	-2,35 (14,77)		0,49 (11,60)		-2,44 (11,65)	
Bez maturity * počet sourozenců		2,95 (11,37)		6,43 (9,82)		5,58 (10,00)
Maturita * počet sourozenců		2,43 (11,19)		5,46 (8,82)		4,98 (9,64)
Vysoká škola * počet sourozenců		3,03 (12,07)		4,36 (8,78)		4,35 (10,36)
KONSTANTA	555,84 (5,32)	462,18 (23,74)	525,21 (4,85)	449,34 (20,72)	561,19 (5,41)	477,46 (21,90)

Zdroj: PISA 2003; autoři.

Poznámka: Počet případů N = 5990, počet škol N = 250.

5. Metodologická diskuze výsledků

Na první pohled by se mohlo zdát, že jsou závěry jednotlivých analytických postupů nekonzistentní. V analýze dat ze žákovského dotazníku nenalezneme významný efekt binárního indikátoru pro existenci sourozenců. Nalezneme zde ale významné interakce mezi existencí sourozenců a typem rodinné struktury a mezi existencí sourozenců a vzděláním rodičů. V analýze úplných dat z rodičovského dotazníku i v analýze mnohonásobně imputovaných dat z rodičovského dotazníku zjistíme, že počet sourozenců má významný negativní efekt na testová skóre a že mezi počtem sourozenců a dalšími proměnnými neexistují významné interakce. Přesto se ale domníváme, že – jak máme v úmyslu ukázat níže – jde

o výsledky navzájem slučitelné a vysvětlitelné. Předtím, než zformulujeme hlavní věcné závěry, bychom nicméně chtěli rozebrat použité metody.

Každá část analýzy má své silnější a slabší stránky. Analýza úplných dat ze žákovského dotazníku je založená na poměrně reprezentativním vzorku, je robustní a relativně snadno zobecnitelná. Závažnou pochybnost však vyvolává způsob měření klíčové vysvětlující proměnné (velikost rodiny) v datech, neboť konstrukce žákovského dotazníku v PISA 2003 umožňuje pouze binární kódování velikosti rodiny (počtu sourozenců). Taková proměnná však nemusí být adekvátním nástrojem pro měření teoretického konceptu počet sourozenců/velikost rodiny. Proto výsledky z analýzy úplných dat z žákovského dotazníku nemusejí být vhodnou odpovědí na vstupní teoretické otázky.

Analýza dat z rodičovského dotazníku nabízí robustnější měřicí nástroje, vykazuje ale problematickou reprezentativitu a zahrnuje analytické volby, na něž zřejmě ani mezi statistiky neexistuje jednoznačný názor. Výsledky analýzy mnohonásobně imputovaných dat a výsledky analýzy dostupných dat jsou nicméně věcně shodné: *s každým dodatečným sourozencem v průměru klesá testové skóre*. Zároveň je lze vyložit způsobem, který je vnitřně logický a konzistentní s výsledky analýzy úplných dat ze žákovského dotazníku. Tento výklad (který shrneme v následujících odstavcích) zahrnuje jak rozdíl v hlavním efektu počtu/existence sourozenců na testová skóre, tak případnou (ne)existenci interakcí. Proto tato zjištění považujeme za hlavní věcné zjištění naší analýzy.

Proč se tedy věcný výsledek založený na modelu s binárním prediktorem (indikujícím existenci společně bydlících sourozenců) a na modelu s imputovanými daty odlišuje? Nejpravděpodobnějším vysvětlením je nepřesnost měření při použití binárního indikátoru, který nepostihuje dostatečně výstižně heterogenitu v populaci žáků, kteří mají alespoň jednoho sourozence. Každý dodatečný sourozenec snižuje v průměru výsledky respondenta. Pro zachycení tohoto efektu je třeba kontinuální proměnná měřící počet sourozenců respondenta.

V datech ze žákovského dotazníku jsme našli interakce počtu sourozenců a dalších kategorizovaných vysvětlujících proměnných, zejména vzdělání rodičů a rodinné struktury. Tyto interakce jsme však nepotvrdili při modelování imputovaných dat, proto se domníváme, že také šlo jen o statistický artefakt špatného měření počtu sourozenců (např. v nevlastních rodinách je obvykle vyšší počet sourozenců, takže proměnná „nevlastní rodina“ do sebe pravděpodobně vstřebala efekt vyššího počtu sourozenců). Máme za to, že se negativní efekt počtu sourozenců (v podobě kontinuální proměnné) vyskytuje ve všech typech rodin a na všech úrovních rodičovského vzdělání ve stejné míře.

Rádi bychom závěrem této metodologické diskuze – s vědomím, že metodologická skepse je v dlouhodobém časovém horizontu předpokladem badatelského pokroku – poukázali na přetrvávající zdroje pochybností o úplnosti a definitivnosti prezentovaných výsledků. V první řadě bychom rádi podtrhli, že použití metody mnohonásobné imputace je spojeno s řadou komplikací: řada dílčích voleb, které jsou nutné pro aplikaci této metody, není jednoznačně vyřešena (to

platí např. pro otázku imputace na shlukově strukturovaných datech). Některé obvyklé analytické nástroje nejsou na imputovaných datech vůbec použitelné (např. některé nástroje pro srovnávání modelů). Podobné nejasnosti existují při kombinované práci s *plausible values* a imputací dalších proměnných.

6. Shrnutí hlavních zjištění a jejich význam

Domníváme se, že hlavní závěr tohoto článku je tento: *počet sourozenců má u českých patnáctiletých žáků významný negativní vliv na testové výsledky*. Tento vliv lze identifikovat u matematické, čtenářské i přírodovědné gramotnosti. Efekt počtu sourozenců na testová skóre se však z velké části realizuje nepřímo, a to prostřednictvím vzdělání rodičů, rodinné struktury, typu studované školy a ročníku. Neidentifikovali jsme podstatnou odchylku od linearity tohoto efektu: každý dodatečný sourozenec má zřejmě stejný negativní vliv na testová skóre. Negativní efekt počtu sourozenců existuje ve všech typech rodin i na všech stupních vzdělanostní kvalifikace rodičů.

Analýza je i nadále s ohledem na použité měřicí nástroje nedokonalá. Data nerozlišují společně bydlící a jinde bydlící sourozence. Podobně nerozlišují vlastní a nevlastní sourozence. Náš způsob měření tedy přehlíží část populační heterogenity, což může vést k tomu, že v modelech vidíme minimální rozdíl v průměrných skórech testů gramotnosti. Zdá se nám, že by tímto směrem mohlo směřovat budoucí výzkumné úsilí.

Domníváme se, že naše analýza má významné implikace pro další stratifikační výzkum v ČR. Analýza ukazuje, že pokus specifikovat modely dosahování statusu bez kontroly velikosti rodiny (počtu sourozenců) může vést ke zkreslení a nadhodnotit sílu efektů vzdělání rodičů a struktury rodiny. Dosavadní česká stratifikačně-výzkumná praxe počet sourozenců jako vysvětlující proměnnou spíše přehlížela. Máme za to, že by tato tradice měla být opuštěna. Pokud se, mimo to, na jiných (a z hlediska kvality měření přesnějších) datech potvrdí existence interakcí mezi počtem sourozenců a typem rodinné struktury a pokud bude narůstat podíl „netradičních“ rodin (osamělý rodič, nevlastní rodič, jiná rodina) v populaci, bude při vynechání počtu sourozenců narůstat riziko špatné specifikace modelu a tedy zkreslení odhadnutých efektů. Proto se domníváme, že je na místě, aby sociologové byli k otázce použití počtu sourozenců jako vysvětlující proměnné ve stratifikačních analýzách otevřenější, než tomu bylo doposud.

MARTA HIRSCHOVÁ studuje magisterské studium sociologie na Fakultě filozofické Západočeské univerzity v Plzni, tématu sourozeneckých konstelací se věnovala ve své magisterské diplomové práci.

MARTIN KREIDL vystudoval postgraduálně sociologii na Kalifornské univerzitě v Los Angeles (M.A. 2001, Ph.D. 2005) a na Fakultě sociálních věd Univerzity Karlovy v Praze (Ph.D. 2003). Zabývá se výzkumem sociálních nerovností a mobility, sociální demografií a kvantitativními aplikacemi v sociálních vědách.

Literatura

- Allison, Paul D. 2000. „Multiple Imputation for Missing Data: A Cautionary Tale.“ *Sociological Methods and Research* 28 (3): 301–309.
- Allison, Paul D. 2002. *Missing Data*. Thousand Oaks: SAGE.
- Becker, G. S. 1964. *Human Capital*. New York: Columbia University Press.
- Becker, G., N. Tomes. 1976. „Child Endowments, and the Quantity and Quality of Children.“ *Journal of Political Economy* 84 (4): 143–162.
- Blake, J. 1981. „Family Size and Quality of Children.“ *Demography* 18 (4): 421–442.
- Blake, J. 1986. „Number of Siblings, Family Background, and the Process of Educational Attainment.“ *Social Biology* 33 (1–2): 5–21.
- Blake, J. 1989. *Family Size and Achievement*. Berkeley: University of California Press.
- Blau, P. M., O. D. Duncan. 1967. *The American Occupational Structure*. New York: John Wiley and Sons.
- Bourdieu, P. 1977. „Cultural Reproduction and Social Reproduction.“ Pp. 487–511 in J. Karabel, A. H. Halsey (eds.). *Power and Ideology in Education*. New York: Oxford University Press.
- Coleman, J. S. 1988. „Social Capital in the Creation of Human Capital.“ *American Journal of Sociology* 94 (1): 95–120.
- Downey, D., S. Neubauer. 1998. „Is Resource Dilution Inevitable? The Association between Number of Siblings and Educational Outcomes across Subgroups.“ Paper presented at the 93rd Annual Meeting of the American Sociological Association. San Francisco.
- Downey et al. 1999. „Much Ado About Siblings: Change Models, Sibship Size and Intellectual Development.“ *American Sociological Review* 64 (2): 193–198.
- Duncan, O. D. 1968. „Ability and Achievement.“ *Eugenics Quarterly* 15 (1): 1–11.
- Eddings, W., J. Marchenko. 2010. „How Can I Account for Clustering When Creating Imputations with MI Impute?“ Stata: Data Analysis and Statistical Software [online; cit. 19. 3. 2012]. Dostupné z: <http://www.stata.com/support/faqs/stat/impute_cluster.html>.
- Enders, C. K. 2010. *Applied Missing Data Analysis*. New York: Guilford Press.
- Ernst, C., J. Angst. 1983. *Birth Order: Its Influence on Personality*. Berlin, New York: Springer-Verlag.
- Featherman, D. L., R. M. Hauser. 1978. *Opportunity and Change*. New York: Academic Press.
- Gailbraith, R. C. 1982. „Sibling Spacing and Intellectual Development: A Closer Look at the Confluence Model.“ *Developmental Psychology* 18 (2): 151–173.
- Guo, G., L. K. VanWey. 1999. „Sibship Size and Intellectual Development: Is the Relationship Causal?“ *American Sociological Review* 64 (2): 169–187.

- Hanushek, E. A. 1992. „The Trade-off between Child Quantity and Quality.“ *Journal of Political Economy* 100 (1): 84–117.
- Hauser, R. M., W. H. Sewell. 1985. „Birth Order and Educational Attainment in Full Sibships.“ *American Educational Research Journal* 22 (1): 1–23.
- Hauser, R. M., W. H. Sewell. 1986. „Family Effects in Simple Models of Education, Occupational Status, and Earnings: Findings from the Wisconsin and Kalamazoo Studies.“ *Journal of Labor Economics* 4 (3): 83–115.
- Kreidl, M. 2004. „Politics and Secondary School Tracking in Socialist Czechoslovakia, 1948–1989.“ *European Sociological Review* 20 (2): 123–139.
- Kreidl, M. 2008. *Cesty ke vzdělání. Vzdělanostní dráhy a vzdělanostní nerovnosti v socialismu*. Plzeň: Nakladatelství ZČU.
- Lu, Y., D. J. Treiman. 2008. „The Effect of Sibship Size on Educational Attainment in China: Period Variations.“ *American Sociological Review* 73 (5): 813–834.
- Maralani, V. 2008. „The Changing Relationship Between Family Size and Educational Attainment Over the Course of Socioeconomic Development: Evidence from Indonesia.“ *Demography* 45 (3): 693–717.
- Mason, W. M. 2001. „Statistical Analysis: Multilevel Methods.“ Pp. 14988–14994 in N. J. Smelser, P. B. Baltes (eds.) *International Encyclopedia of the Social and Behavioral Sciences*. Oxford: Pergamon.
- Park, H. 2008. „Public Policy and the Effect of Sibship Size on Educational Achievement: A Comparative Study of 20 Countries.“ *Social Science Research* 37 (3): 874–887.
- Pfeffermann, D., L. Lavange. 1989. „Regression Models for Stratified Multistage Cluster Samples.“ Pp. 237–260 in C. J. Skinner, D. Holt, T. M. F. Smith (eds.). *Analysis of Complex Surveys*. Chichester: Wiley.
- Phillips, M. 1999. „Sibship Size Academic Achievement: What We Now Know and What We Still Need to Know: Comment on Guo and VanWey.“ *American Sociological Review* 64 (2): 188–192.
- Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Survey*. New York: John Wiley, Sons.
- Shavit, Y., J. L. Pierce. 1991. „Sibship Size and Educational Attainment in Nuclear and Extended Families: Arabs and Jews in Israel.“ *American Sociological Review* 56 (3): 321–330.
- Shavit, Y., H.-P. Blossfeld. 1993. *Persistent Inequality. Changing Educational Attainment in Thirteen Countries*. Boulder: Westview Press.
- Sociologický ústav AV ČR. 2003. *Longitudinální výzkum PISA-L*. Praha: Oddělení Sociologie vzdělání a stratifikace SOÚ AV ČR.
- Steelman, L. C. 1985. „A Tale of Two Variables: The Intellectual Consequences of Sibship Size and Birth Order.“ *Review of Educational Research* 55 (3): 353–386.
- Steelman et al. 2002. „Reconsidering the Effects of Sibling Configuration: Recent Advantages and Challenges.“ *Annual Review of Sociology* 28: 243–269.
- Straková, J. 2010. „Přidaná hodnota studia na víceletých gymnáziích ve světě dostupných datových zdrojů.“ *Sociologický časopis / Czech Sociological Review* 46 (2): 187–210.
- Šafr, J. 2009. „Status Attainment in the Czech Republic after 1990: Role of Cultural and Socioeconomic Resources and Upbringing in the Family of Origin.“ *Working Papers oddělení Studia sociální struktury* 19. Praha: Sociologický ústav AV ČR, v.v.i.
- Ústav pro informace ve vzdělávání (UIV). 2003. *Programme for International Student Assessment (PISA 2003)*. Praha: Oddělení mezinárodních výzkumů.
- Zajonc, R. B., G. B. Markus. 1975. „Birth Order and Intellectual Development.“ *Psychological Review* 82 (1): 74–88.

Příloha

V PISA 2003 tvoří skupinu respondentů patnáctiletí žáci ze základních a středních škol. Celkem bylo dotázáno 9838 žáků. Byla aplikována dvoustupňová metoda výběru. Nejprve bylo vybráno 251 škol ze všech základních a středních škol v České republice, poté byli náhodně vybíráni žáci narození v roce 1987. Datový soubor tvoří odpovědi od 6320 žáků a reprezentuje populaci 121 183 žáků. Soubor vytvořený pro analýzu má po úpravě dat a vynechání případů s chybějícími hodnotami vysvětlujících proměnných 5990 pozorování z 250 škol.

Závislé proměnné představuje testové skóre matematické, čtenářské a přírodovědné gramotnosti. Tato proměnná neobsahuje chybějící hodnoty a je numerického typu. Proměnná přítomnost sourozence v domácnosti obsahovala 134 chybějících pozorování, která byla před analýzou vymazána. Struktura rodiny představuje konstruovanou proměnnou, která nabývá čtyř kategorií: biologičtí rodiče, single rodič, nevlastní rodič, ostatní. Kategorie „ostatní“ obsahuje alternativní typy rodiny a chybějící případy. Vzdělání vzdělanějšího z rodičů je původní proměnná, která byla překódována na kategorie: základní, bez maturity, s maturitou, vysokoškolské vzdělání rodičů. Původní proměnná obsahovala 199 chybějících případů, které byly před analýzou vymazány.

V rámci analýzy je pracováno s 9. a 10. ročníkem¹⁵ (respektive posledním ročníkem primárního stupně a prvním ročníkem sekundárního stupně). Proměnná „ročník“ měla původně 4 kategorie (7., 8., 9. a 10. ročník). Pro účely této analýzy byl vymazán 7. ročník (7 pozorování) a 8. ročník (122 pozorování). Finální proměnná byla dichotomizována. Proměnná „typ školy“ byla upravena do následující podoby: základní škola, víceleté gymnázium, čtyřleté gymnázium, střední odborná škola a střední odborné učiliště s maturitou (SOŠ), střední odborné bez maturity (SOU). Proměnná byla upravena tak, že konzervatoře jsou připojeny k víceletému gymnáziu a zvláštní a praktické školy k základním školám.

¹⁵ 9. ročník označuje jak daný stupeň základní školy, tak jemu odpovídající stupeň gymnaziální (kvarta). Analogicky 10. ročník představuje jak první stupeň střední školy, učňovského oboru, čtyřletého gymnázia či konzervatoře, tak pátý stupeň (kvinta) osmiletého gymnázia či třetí stupeň šestiletého gymnázia.

Tabulka P1. Základní popisné statistiky proměnných vstupujících do analýzy.
Data ze žákovského dotazníku – první část

	N	%	Průměr	SD	Min	Max
Matematická gramotnost						
Pv1math	5990		540,71	95,80	220,94	879,92
Pv2math	5990		541,02	96,45	183,24	849,93
Pv3math	5990		540,38	96,10	183,24	837,70
Pv4math	5990		540,37	96,16	210,73	855,69
Pv5math	5990		541,26	95,80	209,72	865,90
Čtenářská gramotnost						
Pv1read	5990		511,84	90,82	164,79	795,79
Pv2read	5990		511,68	90,70	184,04	793,80
Pv3read	5990		512,23	91,08	147,79	798,17
Pv4read	5990		511,97	90,24	187,25	828,12
Pv5read	5990		513,30	90,77	159,18	788,64
Přírodovědná gramotnost						
Pv1scie	5990		545,95	101,62	172,22	912,72
Pv2scie	5990		546,77	101,78	138,45	874,95
Pv3scie	5990		545,91	101,09	136,64	903,82
Pv4scie	5990		546,68	101,67	95,78	890,48
Pv5scie	5990		548,01	101,23	211,35	911,26
Sourozenci – žákovský dotazník						
Ne	1012	16,89				
Ano	4978	83,11				
Typ rodiny						
Biologičtí rodiče	4548	75,93				
Single rodič	743	12,40				
Nevlastní rodič	644	10,75				
Ostatní	55	0,92				

Tabulka P1. Základní popisné statistiky proměnných vstupujících do analýzy.
Data ze žakovského dotazníku – dokončení

	N	%	Průměr	SD	Min	Max
Nejvyšší vzdělání vzdělanějšího z rodičů						
ZŠ	87	1,45				
Bez maturity	1271	21,22				
Maturita	2753	45,96				
VŠ	1879	31,37				
Typ školy						
ZŠ	2185	36,48				
VG	1158	19,33				
4G	398	6,64				
SOŠ	1529	25,53				
SOU	720	12,02				
Ročník						
9. ročník	2690	44,91				
10. ročník	3300	55,09				

Zdroj: PISA 2003; autoři.

Poznámka: Počet případů N = 5990, počet škol N = 250.

**Tabulka P2. Základní popisné statistiky proměnných vstupujících do analýzy.
Data z rodičovského dotazníku.**

	N	%
Sourozenci – rodičovský dotazník		
0	573	13,88
> 0	3556	86,12
Počet sourozenců – rodičovský dotazník		
0	573	13,88
1	2480	60,06
2	743	17,99
3	174	4,21
4	103	2,49
5	36	0,87
6	13	0,31
7	5	0,12
8	2	0,05
Neodevzdali		
NE	4129	68,93
ANO (chybějící pozorování)	1861	31,07

Zdroj: PISA 2003; autoři.

Poznámka: Počet úplných případů N=4129, počet škol N=243.