

- úroveň 2: domácnosti s právě jedním dospělým, který není součástí ekonomicky aktivní populace;
- úroveň 3: domácnosti, kde jsou všichni dospělí součástí ekonomicky aktivní populace.
- Ordinální proměnná „horní segment“, reprezentující podíl osob vykonávajících povolání v horním segmentu terciárního sektoru z celkového počtu dospělých členů domácnosti, kteří jsou součástí ekonomicky aktivní populace. Jsou zde tři úrovně:
 - úroveň 1: domácnosti, kde méně než 40 % dospělých členů, kteří jsou součástí ekonomicky aktivní populace, vykonává povolání v horním segmentu terciárního sektoru;
 - úroveň 2: domácnosti, kde 40 až 60 % dospělých členů, kteří jsou součástí ekonomicky aktivní populace, vykonává povolání v horním segmentu terciárního sektoru;
 - úroveň 3: domácnosti, kde více než 60 % dospělých členů, kteří jsou součástí ekonomicky aktivní populace, vykonává povolání v horním segmentu terciárního sektoru.
- Ordinální proměnná „bohatá země“ reprezentuje podíl osob v rodině, které se narodily v bohaté zemi včetně Belgie. Jsou zde tři úrovně:
 - úroveň 1: domácnosti, kde se méně než 40 % členů narodilo v bohaté zemi;
 - úroveň 2: domácnosti, kde se 40 až 60 % členů narodilo v bohaté zemi;
 - úroveň 3: domácnosti, kde se více než 60 % členů narodilo v bohaté zemi.
- Nominální proměnná „rodina“ rozlišuje rodiny s jedním rodičem a ty ostatní. Hodnota 1 je připsána rodinám s jedním rodičem a hodnota 2 ostatním.
- Nominální proměnná „CPAS“ (*Centre public d'aide sociale*) rozlišuje rodiny, kde je alespoň jeden člen příjemcem dávek na sociální integraci (hodnota 1), a ty ostatní (hodnota 2).

Jednorozměrnost škály byla testována pomocí Cronbachova alfa. Z celkové populace (847 406 žáků) bylo 708 071 žáků, u nichž nechyběla žádná data (16 % žáků mělo tedy ve svých datech nejméně jednu chybějící informaci). V souboru žáků bez chybějících dat bylo Cronbachovo alfa 0,63.

Protože se v tomto modelu socioekonomický index počítá pro každého jednotlivého žáka, je nezbytné vypořádat se s problémem chybějících dat. Pro přiřazení platných hodnot žákům s chybějícími daty existuje několik metod. Nejjednodušší způsob spočívá v tom, že se jim připíše průměr dané proměnné, přičemž se předpokládá, že osoba s chybějícím údajem může být připodobněna k průměrnému jedinci. Tato technika ovšem vede ke značnému zkreslení v odhadu parametrů populace. Další techniky založené na modelech mnohonásobné imputace dávají lepší statistické záruky, avšak předpokládají vícerozměrné normální rozdělení spojitých proměnných, zatímco některé z našich proměnných jsou diskrétní (buď kategorické, jako je úroveň dosaženého vzdělání, anebo nominální, jako třeba proměnná rodina).

Aby se vypořádal s chybějícími hodnotami, rozhodl se výzkumný tým použít modely teorie odpovědi na položku (IRT). Tyto modely dávají chybějícím datům psychometrický analytický rámec, který nepředpokládá žádný konkrétní typ nebo rozdělení dat. Umožňují odhadnout individuální hodnoty (v našem případě individuální socioekonomický index) ze sady proměnných, které se kvůli chybějícím datům mohou u jednotlivých žáků lišit. Výsledný socioekonomický index jednotlivých žáků se však dá navzájem porovnávat, a to i tam, kde jsou chybějící hodnoty v některých vstupních proměnných. Tabulka č. 8 uvádí pro 847 899 žáků

