



Methodologische document  
Leuvens Adolescenten- en Gezinnenonderzoek  
Versie 2.0

Sofie Vanassche - An Katrien Sodermans – Graziela Dekeyser  
Koen Matthijs

Onderzoeksverslag  
Centrum voor Sociologisch Onderzoek (CeSO)

**Onderzoeksdomein Family and Population Studies**

CeSO/GB/2011-3

# Inhoud

## *DEEL 1: Steekproefopzet en -kenmerken*

1. Beschrijving van de steekproefopzet .....	3
1.1. Fase 1: Selectie van scholen .....	3
1.2. Fase 2: Selectie van leerlingen binnen scholen .....	5
2. LAGO binnen het secundaire onderwijslandschap in Vlaanderen .....	6
2.1. Schoolkenmerken voltijds gewoon secundair onderwijs .....	6
2.2. Leerlingenkenmerken voltijds gewoon secundair onderwijs .....	8
3. Besluit.....	11

## *DEEL 2: Datacleaning*

1. Datacleaning op itemniveau (door projectcoördinator) .....	13
2. Datacleaning op respondentenniveau (door projectcoördinator) .....	13
3. Datacleaning modules B en C (door eindgebruiker) .....	14

# DEEL 1

## Steekproefopzet en –kenmerken

### 1. Beschrijving van de steekproefopzet

Binnen het Leuvens Adolescenten- en Gezinnenonderzoek (LAGO) wordt gewerkt met een tweetrapssteekproef, waarbij in een eerste fase scholen worden geselecteerd en in een tweede fase klassen (van leerlingen) binnen deze scholen. In de volgende paragrafen worden deze twee fasen uitgebreid besproken.

#### 1.1. Fase 1: Selectie van scholen

De selectie van de scholen gebeurt via een disproportionele quota-steekproef. Het betreft een quotasteekproef omdat de scholenpopulatie vooraf in een aantal subgroepen of quota verdeeld worden op basis van onderwijsnet, provincie en gemeente. Uit elk van deze subgroepen worden vervolgens scholen geselecteerd. Het betreft een disproportionele steekproef omdat er bij de samenstelling van de scholensteekproef niet zozeer gestreefd wordt naar een identieke procentuele verhouding van de subgroepen zoals zij voorkomen in de Vlaamse scholenpopulatie, maar wel naar een voldoende aantal eenheden binnen alle subgroepen. Post hoc gaan we wel kijken in welke mate de uiteindelijke verdeling van de subgroepen in onze scholensample overeenkomt met die van de Vlaamse scholen. Zodoende kunnen we voor grote afwijkingen corrigeren via weegcoëfficiënten.

De doelstelling is om in vooraf afgebakende geografische regio's binnen elk onderwijsnet en elke studierichting scholen op te nemen in de steekproef. Dit gebeurt in twee fasen. In een eerste fase worden de 'eigen' scholen van de meewerkende studenten gecontacteerd en andere scholen waarmee de studenten of verwanten een persoonlijke band hebben. De motivatie achter deze werkwijze is dat persoonlijke contacten met de school de responsgraad sterk verhogen. Dit is cruciaal voor een onderzoek met beperkte middelen en tijd. Alle scholen die in deze fase werden gecontacteerd gaven toestemming om mee te werken. In totaal werden zo over de vier bevragingronde heen 19 scholen gecontacteerd, allen uit het gesubsidieerd vrij onderwijsnet. Tien van deze scholen liggen in de provincie Antwerpen, drie in Vlaams-Brabant, vijf in Limburg en één in West-Vlaanderen. Een nadeel bij deze werkwijze is dat de masterstudenten die de opleiding sociologie aan de KU Leuven volgen een specifiek profiel vertonen. Dit leidt tot een overselectie van ASO-scholen, scholen uit het vrije onderwijsnet en uit de provincies Vlaams-Brabant, Limburg en Antwerpen. In een tweede fase wordt deze overselectie gecompenseerd via een aselechte, disproportioneel gestratificeerde steekproef.

Tabel 1: Gerealiseerd en beoogd aantal scholen voor alle strata, LAGO-steekproef

Net	Stratum		Aanwezig in ronde	Gerealiseerd aantal scholen			Quotum
	Provincie	Gemeente		Fase 1	Fase 2	Totaal	
GO(O)	Antwerpen	Stad Antwerpen	1	0	1	1	2
		Mechelen	2	0	3	3	2
		Overige	2	0	3	3	2
	Limburg	Hasselt	1 & 3	0	0	0	2
		Genk	1 & 3	0	0	0	2
		Overige	1 & 3	0	0	0	2
	Vlaams- Brabant	Leuven	1 & 3	0	1	1	2
		Vlaamse Rand Brussel	2	0	2	2	2
		Overige	2	0	1	1	2
	West- Vlaanderen	Brugge	3	0	1	1	2
		Oostende	3	0	0	0	2
		Kortrijk	3	0	1	1	2
		Overige	3	0	0	0	2
	Oost- Vlaanderen	Gent	3	0	0	0	2
		Aalst	4	0	1	1	2
		Sint-Niklaas	-	0	-	-	2
		Overige	3	0	0	0	2
	Brussels Hoofdstedelijk Gewest	Stad Brussel	-	-	-	-	2
		Anderlecht	4	0	0	0	2
		Sint-Jan-Molenbeek	-	-	-	-	2
Overige		4	0	1	1	2	
GVO	Antwerpen	Stad Antwerpen	1	1	0	1	2
		Mechelen	2	0	0	0	2
		Overige	1, 2 & 3	9	0	9	2
	Limburg	Hasselt	1 & 3	0	1	1	2
		Genk	1 & 3	0	0	0	2
		Overige	1, 3 & 4	5	1	6	2
	Vlaams-Brabant	Leuven	1	1	0	1	2
		Vlaamse Rand Brussel	2	0	0	0	2
		Overige	2	2	0	2	2
	West- Vlaanderen	Brugge	3	0	1	1	2
		Oostende	3	0	1	1	2
		Kortrijk	3,4	1	1	2	2
		Overige	3	0	0	0	2
	Oost- Vlaanderen	Gent	3	0	1	1	2
		Aalst	4	0	2	2	2
		Sint-Niklaas	-	-	-	-	2
		Overige	3	0	1	1	2
	Brussels Hoofdstedelijk Gewest	Stad Brussel	-	-	-	-	2
		Anderlecht	4	0	1	1	2
		Sint-Jan-Molenbeek	-	-	-	-	2
Overige		4	0	0	0	2	
TOTAAL				19	25	44	84

Tabel 1 bevat een overzicht van de verschillende strata, gebaseerd op specifieke combinaties van onderwijsnet, provincie en gemeente. Alle strata hebben betrekking op geografische regio's van minstens 50.000 inwoners. De eerste kolom geeft voor alle strata weer in welke bevragingronde ze werden opgenomen. De twee volgende kolommen tonen per stratum het aantal scholen dat in de steekproef aanwezig is, afzonderlijk voor fase 1 (eigen of gekende scholen) en fase 2 (niet-gekende scholen). De vierde kolom geeft het totaal aantal scholen weer voor de twee fasen samen. De laatste kolom geeft voor alle strata het quotum weer, m.a.w. het beoogde aantal scholen dat we in de steekproef wensen op te nemen.

De strata hebben als doelstelling 1) de scholen geografisch te clusteren per bevragingronde (omwille van praktische redenen) met behoud van voldoende geografische spreiding over de volledige sample en 2) een voldoende grote, d.i. onderzoeksvatbare, proportie van scholen uit de verschillende onderwijsnetten en regio's te garanderen. Binnen de strata van stad Antwerpen, Gent, Genk, Mechelen, de Vlaamse Rand van Brussel en het Brussels Hoofdstedelijk Gewest verwachten we bovendien een onderzoeksvatbaar aandeel van de Vlaamse migrantenpopulatie terug te vinden. Door steden te selecteren met een grote migrantenpopulatie wordt immers de kans op een school met een groot aandeel migranten in hun leerlingenpopulatie verhoogd. Binnen elk stratum wordt de voorkeur gegeven aan scholen waarin verschillende onderwijsvormen worden aangeboden. Scholen met verschillende onderwijsvormen krijgen hiervoor een verhoogd gewicht in de random schoolselectie per stratum. Deze werkwijze heeft als doel om het overgewicht aan ASO-scholen uit de eerste fase te compenseren en om binnen elk onderwijsnet alle onderwijsvormen op te nemen.

## **1.2. Fase 2: Selectie van leerlingen binnen scholen**

Binnen de geselecteerde scholen wordt vervolgens een clustersteekproef genomen van klassen waarin alle leerlingen worden bevroegd. De doelstelling is om per graad en per onderwijsvorm (ASO, KSO, TSO en BSO) twee klassen te bevragen. Dit betekent dat er bijvoorbeeld in een school met drie graden en drie onderwijsvormen beoogd wordt om achttien klassen te bevragen. Een bijkomend criterium is dat binnen graden en onderwijsvormen met typische jongens- en meisjesrichtingen (bv. automechanica, kinderverzorging), van elk van beide een klas wordt geselecteerd. De selectie van de klassen gebeurt telkens in overleg met de directie, waarbij zoveel mogelijk de willekeurigheid van selectie wordt nagestreefd. Om te voldoen aan de voorwaarden tot deelname die de schooldirecties voorop stellen, dienen verschillende procedures van vragenlijstafname te worden gevolgd. Sommige scholen laten de vragenlijsten klassikaal afnemen in overleg met bijvoorbeeld de leerkracht godsdienst. Anderen laten de vragenlijsten afnemen tijdens ziekte van een leerkracht of tijdens een vrij studie-uur.

Detailanalyses van de bevroegde leerlingen per school leren dat de richtlijn van twee klassen per graad en onderwijsvorm effectief werd toegepast: in iedere school zijn voor iedere aangeboden graad en onderwijsvorm leerlingen bevroegd. Het aantal bevroegde leerlingen per graad en onderwijsvorm varieert sterk, wat in overeenstemming is met de werkelijke variatie in klasgrootte.

## **2. LAGO binnen het secundaire onderwijslandschap in Vlaanderen**

Het is belangrijk om na te gaan in welke mate de bevroegde jongeren uit de LAGO-steekproef de Vlaamse middelbare leerlingenpopulatie weerspiegelen. Op de website onderwijsstatistieken van de Vlaamse Overheid (<http://www.ond.vlaanderen.be/onderwijsstatistieken/>) is uitgebreid informatie terug te vinden over de kenmerken van de leerlingenpopulatie in het secundair onderwijs. We maken gebruik van de gegevens uit het statistisch jaarrapport 2009-2010 en van het scholenoverzicht van 2010-2011. Vergelijkingen met gegevens uit vorige jaren leren dat de verdeling van de hieronder besproken kenmerken relatief stabiel zijn over de vier afgelopen jaren waarin de leerlingen werden bevroegd.

In 2010-2011 stonden officieel 960 secundaire onderwijsinstellingen geregistreerd in Vlaanderen. Een groot deel scholen is echter opgedeeld in verschillende deelinstanties waarbij de eerste graad of 'middenschool', en de bovenbouw of specifieke onderwijsvormen als afzonderlijke onderwijsinstellingen staan geregistreerd. Als we deze verschillende deelinstanties als één school beschouwen (onder voorwaarde dat ze op hetzelfde adres of naast elkaar zijn gevestigd), telt Vlaanderen 706 secundaire onderwijsinstellingen. Tijdens de vier bevroegingsrondes namen reeds 44 of ongeveer 5% van deze scholen deel aan het onderzoek: 10 scholen in de eerste ronde, 10 scholen in de tweede ronde, 15 scholen in de derde ronde en 9 scholen in de vierde ronde.

In totaal waren tijdens het schooljaar 2009-2010 448.760 leerlingen ingeschreven in het secundair onderwijs, waarvan 429.745 in het gewoon secundair onderwijs en 19.015 in het buitengewoon secundair onderwijs. Omdat in LAGO geen leerlingen uit het buitengewoon secundair onderwijs worden bevroegd laten we deze laatste groep ook bij het vergelijkend Vlaamse cijfermateriaal buiten beschouwing. De vier eerste bevroegingsrondes van LAGO resulteerden in 7035 respondenten, wat correspondeert met 1.6% of meer dan 1 op 100 van alle Vlaamse jongeren in het gewoon secundair onderwijs. In de eerste ronde werden 1970 jongeren bevroegd, in de tweede ronde 1688, in de derde ronde 2120 en in de vierde ronde 1257.

We maken in de volgende paragrafen een onderscheid tussen enerzijds schoolkenmerken en anderzijds leerlingenkenmerken. Bij de schoolkenmerken bekijken we de verdeling van de Vlaamse scholen en leerlingen over de verschillende provincies en onderwijsnetten. Bij de leerlingenkenmerken bekijken we de verdeling naar geslacht, studiegraad, onderwijsvorm en nationaliteit.

### **2.1. Schoolkenmerken voltijds gewoon secundair onderwijs**

Schoolkenmerken kunnen zowel op niveau van de schoolpopulatie als van de leerlingenpopulatie worden bekeken. In Tabel 2 wordt de spreiding van de scholen en leerlingen over de verschillende provincies en onderwijsnetten binnen de LAGO-steekproef vergeleken met die binnen de Vlaamse populatie. Zowel op schoolniveau als op leerlingenniveau zien we in de LAGO-steekproef een lichte oververtegenwoordiging van het gemeenschapsonderwijs (GO) ten opzichte van het gesubsidieerd vrij onderwijs (GVO). Onder het gesubsidieerd officieel onderwijs (GOO) vallen de gemeentelijke en provinciale onderwijsinstellingen en die van de Vlaamse overheid. Dit aandeel is overal relatief klein. Daarnaast zien we in de LAGO-steekproef een

oververtegenwoordiging van de provincies Antwerpen en Limburg en een ondervertegenwoordiging van de provincies Oost- en West-Vlaanderen.

Tabel 2: De Vlaamse en LAGO-leerlingenpopulatie naar onderwijsnet en provincie (%)

	Vlaanderen		LAGO	
	Scholen (N=960)	Leerlingen (N=429.745)	Scholen (N=44)	Leerlingen (N=7.035)
<b>Onderwijsnet</b>				
Gemeenschapsonderwijs	23	17	27	25
Gesubsidieerd vrij onderwijs	70	75	66	67
Gesubsidieerd officieel onderwijs	7	8	7	8
<b>Provincie (en BHG)</b>				
Antwerpen	27	28	39	41
Vlaams-Brabant	13	14	16	16
Brussels Hoofdstedelijk Gewest	4	3	4	1
West-Vlaanderen	22	19	14	11
Oost-Vlaanderen	20	22	11	9
Limburg	14	14	16	22

Bronnen: <http://www.ond.vlaanderen.be/onderwijsstatistieken/> & LAGO, eigen berekeningen

Om een idee te krijgen van de impact van deze afwijkende verdelingen op onze steekproefresultaten, bekijken we de samenhang van het onderwijsnet en de provincie met de onderwijsvorm en nationaliteit van de leerlingen. Tussen de verschillende onderwijsnetten is er vooreerst een groot verschil in de verdeling van de onderwijsvorm van de leerlingen (Tabel 3). Het aandeel ASO-leerlingen is het grootst in het GVO, het aandeel TSO-, KSO- en BSO-leerlingen in het GOO. Het aandeel ASO-leerlingen is het kleinst in het GOO, het aandeel TSO-leerlingen is het kleinst in het GO en het aandeel KSO- en BSO-leerlingen is het kleinst in het GVO. Deze verhoudingen zijn binnen de LAGO-steekproef in grote lijnen gelijklopend, maar zeker niet identiek. Er is binnen LAGO een oververtegenwoordiging van ASO-leerlingen in het GVO en van KSO-leerlingen binnen het GOO, en een ondervertegenwoordiging van BSO-leerlingen binnen het GOO en het GVO.

**Tabel 3: De verschillende onderwijsvormen (2de en 3de graad) binnen de verschillende onderwijsnetten in Vlaanderen en de LAGO-steekproef (%)**

Onderwijsnet	Vlaanderen (N=290.517)			
	ASO	TSO	KSO	BSO
Gemeenschapsonderwijs	38	24	2	36
Gesubsidieerd vrij onderwijs	43	33	1	22
Gesubsidieerd officieel onderwijs	12	36	8	44

  

Onderwijsnet	LAGO (N=4.803)			
	ASO	TSO	KSO	BSO
Gemeenschapsonderwijs	37	28	2	33
Gesubsidieerd vrij onderwijs	56	29	2	13
Gesubsidieerd officieel onderwijs	21	39	19	21

Bronnen: <http://www.ond.vlaanderen.be/onderwijsstatistieken/> & LAGO, eigen berekeningen

Tussen de verschillende onderwijsnetten is er ook een groot verschil in het aandeel leerlingen van een vreemde nationaliteit (Tabel 4). Het aandeel leerlingen met een vreemde nationaliteit is dubbel zo groot in het officieel onderwijs dan in het vrije onderwijsnet. Daarenboven zien we ook een groot verschil in het aandeel niet-Belgen tussen de verschillende provincies, waarbij het aandeel in Limburg vijf keer zo groot is als in West-Vlaanderen. Deze verhoudingen zijn relatief gelijklopend in de LAGO-steekproef, met die uitzondering van een hoger aandeel niet-Belgen binnen Vlaams-Brabant. Dit komt door de oververtegenwoordiging van de Vlaamse rand rond Brussel in de steekproef.

**Tabel 4: Leerlingen met vreemde nationaliteit binnen de verschillende onderwijsnetten en provincies in Vlaanderen en binnen de LAGO-steekproef (%)**

Onderwijsnet	Aandeel niet-Belgen	
	Vlaanderen (N= 429.745)	LAGO (N=6.981)
Gemeenschapsonderwijs	8	10
Gesubsidieerd vrij onderwijs	4	6
Gesubsidieerd officieel onderwijs	10	9

  

Provincie	Vlaanderen (N= 429.745)	LAGO (N=6.981)
Antwerpen	7	7
Vlaams-Brabant	3	7
Brussels Hoofdstedelijk Gewest	9	19
West-Vlaanderen	2	7
Oost-Vlaanderen	4	3
Limburg	10	9

Bronnen: <http://www.ond.vlaanderen.be/onderwijsstatistieken/> & LAGO, eigen berekeningen



Onderwijsnet en provincie hangen ook onderling samen. Het aandeel van de verschillende onderwijsnetten is relatief gelijklopend in de verschillende provincies, met die uitzonderingen dat het Brussels Hoofdstedelijk Gewest een relatief laag aandeel scholen uit het GVO kent (64%) en West-Vlaanderen een relatief hoog aandeel (84%). In de andere provincies cirkelt de verdeling vrij versus officieel onderwijs op leerlingniveau rond 75% versus 25%.

## 2.2. Leerlingenkenmerken voltijds gewoon secundair onderwijs

In Tabel 5 zien we de verdeling naar geslacht, onderwijsvorm en –graad voor zowel de Vlaamse als bevroegde scholierenpopulatie. Over het algemeen kunnen we stellen dat de gezamenlijke verdeling van deze drie kenmerken in de bevroegde populatie heel sterk die van de Vlaamse populatie benadert. In de LAGO-onderzoekspopulatie is er een lichte oververtegenwoordiging van meisjes, van ASO-leerlingen versus BSO-leerlingen en van leerlingen uit de tweede en derde graad versus de eerste graad. Deze disproportionaliteit kan worden opgevangen met behulp van weegcoëfficiënten. Alle subgroepen zijn ook voldoende groot om betekenisvolle uitspraken te kunnen doen omtrent jongens-meisjes verschillen, verschillen tussen onderwijsvormen en leeftijdsverschillen (studiejaar).

Tabel 5: Verdeling (%) van de Vlaamse en bevroegde populatie naar geslacht, studiegraad en onderwijsvorm

Graad	Onderwijsvorm	Vlaanderen (N=425.316)			LAGO (N=6.919)		
		Jongens	Meisjes	Totaal	Jongens	Meisjes	Totaal
Eerste	1ste leerjaar A en 2de leerjaar	13.1	13.4	26.5	11.0	11.6	22.6
	1ste leerjaar B en beroepsvoorbereidend leerjaar	3.2	2.5	5.7	1.5	2.1	3.6
	<i>Totaal eerste graad</i>	16.3	15.9	32.2	12.5	13.7	26.2
Tweede	ASO	6.7	7.9	14.6	9.3	9.9	19.2
	TSO	5.8	4.3	10.1	6.1	5.2	11.3
	KSO	0.2	0.4	0.7	0.4	0.9	1.3
	BSO	4.0	3.3	7.3	2.6	4.2	6.7
	<i>Totaal tweede graad</i>	16.7	15.9	32.6	18.3	20.2	38.6
Derde & vierde	ASO	5.6	7.1	12.7	6.2	9.0	15.2
	TSO	6.5	5.1	11.7	4.9	5.2	10.0
	KSO	0.3	0.5	0.8	0.3	0.8	1.0
	BSO	5.4	4.7	10.1	3.6	5.4	9.0
	<i>Totaal derde &amp; vierde graad</i>	17.8	17.4	35.2	14.9	20.3	35.3
<i>Totaal</i>		50.8	49.1	100.0	45.7	54.3	100.0

Bronnen: <http://www.ond.vlaanderen.be/onderwijsstatistieken/> & LAGO, eigen berekeningen

In Tabel 6 worden een aantal additionele leerlingenkenmerken vergeleken voor de bevroagde en Vlaamse scholierenpopulatie. Een eerste kenmerk betreft de gezinssituatie, een centraal concept in het LAGO-project. Het aandeel bevroagde jongeren dat een ouderlijke scheiding meemaakte is perfect gelijklopend met het geschatte aandeel Vlaamse jongeren tussen 12 en 17 jaar met gescheiden ouders (Lodewijckx, 2005)<sup>1</sup>.

De kenmerken *opleidingsniveau moeder* en *thuis taal* voor de Vlaamse populatie zijn gebaseerd op de GOK-indicatoren. GOK staat voor Gelijke Onderwijskansen. Deze duiden aan of 1) de moeder al dan niet een diploma hoger secundair onderwijs heeft, en 2) de taal in het gezin niet de onderwijstaal is, wat wil zeggen dat niemand in het gezin de onderwijstaal spreekt, of in een gezin van drie gezinsleden maximum één gezinslid de onderwijstaal spreekt (broers en zussen tellen als 1 gezinslid). Bij het vergelijken van de cijfers voor opleidingsniveau is enige voorzichtigheid geboden daar het LAGO-onderzoek dit bevroagt bij de jongeren zelf, dus niet bij de ouders. Hierbij werd geen onderscheid gemaakt tussen een diploma lager en hoger secundair onderwijs. De minimum (8%) en maximumschatting (43%) van het aandeel jongeren met laag opgeleide moeders in LAGO wijzen echter op een gelijkaardige verdeling als in de Vlaamse scholierenpopulatie. De thuis(a)l(en) werd(en) niet bevroagd in de eerste bevroagingsronde en op een verschillende manier in de overige bevroagingsrondes. De operationalisatie voor de laatste twee bevroagingsrondes is identiek aan die voor de GOK -indicator en het percentage ligt ook in dezelfde lijn. Het aandeel jongeren met een andere thuis taal in de tweede bevroagingsronde houdt geen rekening met de personen met wie de talen worden gesproken. Er wordt in dit geval gesproken over een andere thuis taal indien de leerling geen Nederlands aanduidde of naast het Nederlands nog twee andere talen. Het percentage met een andere thuis taal in de tweede ronde is hoger dan in de Vlaamse populatie. Dit kan deels een verschil in operationalisatie betreffen, maar zal zeker ook deels een gevolg zijn van de steekproefsamenstelling in de tweede ronde met een hoger aandeel scholen uit de officiële onderwijsnetten en uit steden met een grote migrantenpopulatie.

Het aandeel jongeren met de niet-Belgische nationaliteit ligt tenslotte iets hoger in de bevroagde populatie. Toch kunnen we stellen dat het aandeel jongeren van Belgische nationaliteit in de LAGO-steekproef (92%) niet problematisch afwijkt van het aandeel in de Vlaamse populatie (95%). Voor de cijfers inzake herkomst is geen vergelijkend cijfermateriaal voor de Vlaamse populatie gevonden. Binnen LAGO zien we dat (in de tweede, derde en vierde ronde) ongeveer 30% van de jongeren minstens één ouder of grootouder heeft die niet in België werd geboren of er zelf niet geboren is.

---

<sup>1</sup> Lodewijckx, E. (2005). *Kinderen en scheiding bij hun ouders in het Vlaamse Gewest. Een analyse op basis van Rijksregistergegevens*. Brussel, CBGS, Werkdocument 7.

**Tabel 6: Jongeren met gescheiden ouders, een laag opgeleide moeder, een andere thuistaal, een vreemde nationaliteit en van vreemde herkomst in de Vlaamse en bevroagde scholierenpopulatie**

	Vlaamse leerlingen (N=429.745)	Bevroagde leerlingen (N=5.778)
<b>Met gescheiden ouders</b>	26%	26% (Exclusief Iln met overleden ouder)
<b>Met laagopgeleide moeder</b>	25% zonder diploma hoger secundair onderwijs	8% enkel diploma lager onderwijs 35% enkel diploma secundair onderwijs
<b>Met andere thuistaal dan het Nederlands</b>	10% 8% zonder Brussels Hoofdstedelijk Gewest	16% (Ronde 2) 9% (Ronde 3 & 4)
<b>Met niet-Belgische nationaliteit</b>	5%	7%
<b>Herkomst</b>	<i>(Geen informatie gevonden)</i>	(Ronde 2, 3 & 4)
niet in België geboren		7%
minstens één ouder in ander land geboren		16%
minstens één grootouder in ander land geboren		6%
respondent, ouders en grootouders in België geboren		71%

Bronnen: Lodewijckx, E. (2005). Kinderen en scheiding bij hun ouders in het Vlaamse Gewest. Een analyse op basis van Rijksregistergegevens. Brussel, CBGS, Werkdocument 7, <http://www.ond.vlaanderen.be/onderwijsstatistieken/> & LAGO, eigen berekeningen

Binnen de groep niet-Belgen of jongeren van vreemde herkomst is natuurlijk een grote variatie in de landen van herkomst. Op dit vlak vertoont de verdeling tussen de Vlaamse leerlingpopulatie en de bevroagde leerlingpopulatie belangrijke verschillen. De LAGO-steekproef bevat bijvoorbeeld slechts voor een beperkt aantal herkomstlanden een voldoende groot aantal eenheden om uitspraken over te doen. Aan onderzoekers die de LAGO-data willen gebruiken voor migrantenonderzoek wordt aangeraden om in een eerste fase een grondige verkenning te doen van de samenstelling en kenmerken van de migranten binnen de sample.

### 3. Besluit

Quota-steekproeven worden soms bekritiseerd omwille van hun niet aselekt karakter. Een toevalssteekproef, die wel aselekt getrokken wordt, garandeert echter geen perfecte representativiteit, maar is slechts een hulpmiddel om dit te bereiken. Een representatieve steekproef weerspiegelt de empirische populatie goed op de relevante kenmerken en in het ideale geval op de gezamenlijke verdeling van deze kenmerken. In het algemeen kunnen we stellen dat de LAGO-scholensteekproef niet volledig ad-random is en niet representatief voor de scholenpopulatie naar onderwijsnet en geografische ligging (provincies/gemeentes). Deze scholensteekproef is echter slechts een tussenstap in de leerlingselectie. De samenstelling van de leerlingsteekproef over de vier bevroagingsrondes is een mooie afspiegeling van de

Vlaamse middelbare scholierenpopulatie en dit zowel voor het schoolkenmerk onderwijsnet, als voor de gezamenlijke verdeling op de leerlingenkenmerken geslacht, studiegraad en onderwijsvorm (ASO, TSO, KSO en BSO) en het aandeel jongeren met gescheiden ouders (de inhoudelijke focus van het project). De data zijn met andere woorden voldoende representatief om betekenisvolle uitspraken te kunnen doen over de Vlaamse jeugd. Voor het veralgemenen van uitspraken over geslacht, studiejaar en onderwijsvormen kunnen de kleine afwijkingen worden opgevangen via weegcoëfficiënten. Waar tot op heden geen veralgemeende uitspraken over kunnen worden gedaan met het huidige LAGO-bestand zijn verschillen tussen provincies en gemeentes. Door de disproportionele vertegenwoordiging van de verschillende provincies binnen de onderwijsnetten is ook voorzichtigheid geboden bij de interpretatie van verschillen tussen de onderwijsnetten: hierbij dient steeds gecontroleerd te worden voor regionale verschillen.

Door de wijze van steekproeftrekking en de daaruit volgende datastructuur wordt idealiter een multilevel techniek gebruikt voor de analyse van de data. De leerlingen zitten immers geclusterd binnen scholen (en klassen), en zijn dus niet onafhankelijk van elkaar. Daarenboven kunnen in een multilevel-model het effect van school- en omgevingskenmerken worden geïntroduceerd.

Extra voorzichtigheid is geboden wanneer wordt gewerkt met data van slechts één bevragsingsronde, daar de samenstelling van de steekproef over de verschillende provincies en onderwijsnetten stapsgewijs gebeurt over de verschillende bevragsingsrondes. Het is de verantwoordelijkheid van de individuele onderzoeker om zelf altijd voldoende kennis te nemen van de onderzoekspopulatie die hij/zij analyseert.

De migrantensample in LAGO dient te worden beschouwd als een exploratieve sample. Omdat er geen officiële cijfers beschikbaar zijn omtrent de samenstelling van scholen naar nationaliteit of herkomst van de leerlingen is het onmogelijk om bij de steekproeftrekking deze parameter onder controle te houden. Een bruikbare proxy om de socio-culturele samenstelling van de scholen in kaart te brengen zijn de GOK-indicatoren 'andere thuistaal' en 'opleidingsniveau moeder'.

## DEEL 2

# Datacleaning Leuvens Adolescenten- en Gezinnenonderzoek

Om van een ruwe dataset tot betrouwbaar analyse materiaal te komen is een grondige datacleaning nodig. Deze cleaning gebeurt in drie fasen. De eerste twee fasen worden door de projectcoördinator uitgevoerd, de laatste dient door de eindgebruiker zelf te worden uitgevoerd.

### 1. Datacleaning op itemniveau (door projectcoördinator)

Een eerste fase is de datacleaning op itemniveau waarbij alle waarden in strijd met het codeboek op missing worden geplaatst. Dit zijn immers ongeldige waarden. Het gaat hierbij om:

- Ongeldige antwoorden door respondenten, ingegeven door codeurs, bv. de respondent plaatste een kruisje tussen twee antwoordcategorieën, wat door een codeur foutief als het gemiddelde van de twee antwoordcategorieën werd ingegeven.
- Fouten tijdens het ingeven van de vragenlijsten, bv. een onbestaande code.
- Onmogelijke combinaties op basisvariabelen:
  - o De leeftijd bij scheiding van de ouders is groter dan de huidige leeftijd.
  - o De leeftijd bij het samenwonen met de nieuwe partner van moeder/vader is groter dan de huidige leeftijd.
  - o De leeftijd bij een verandering in de verblijfsregeling na scheiding is groter dan de huidige leeftijd.

Deze fase van de cleaning heeft geen invloed op het aantal onderzoekseenheden in de dataset, maar vermindert het aantal observaties voor specifieke variabelen. Het aantal ongeldige waarden over alle variabelen heen is echter heel beperkt.

### 2. Datacleaning op respondentenniveau (door projectcoördinator)

Een volgende fase is de datacleaning op respondentenniveau. In deze fase worden volledige datalijnen uit de dataset verwijderd omwille van verschillende redenen.

In een eerste stap worden alle datalijnen verwijderd waarvoor geen informatie beschikbaar is over geslacht, leeftijd of burgerlijke staat van de ouders. Deze informatie is immers cruciaal om enige uitspraak te doen over de respondent in kwestie. Hierdoor werden 78 datalijnen verwijderd. Deze datalijnen hadden vaak een groot aantal missings over alle variabelen heen.

Daarna worden alle datalijnen verwijderd waarin meer dan een derde (33%) van de basisvariabelen (=variabelen aanwezig in de vier bevraging rondes) missing zijn. Hierdoor werden nog 107 bijkomende datalijnen verwijderd. Dit hing opnieuw sterk samen met een hoge algemene item non-respons.

Vervolgens worden datalijnen van verdachte betrouwbaarheid verwijderd op basis van de opmerkingen van codeurs. De criteria de hiervoor worden gebruikt zijn:

- Terugkerende systematiek in antwoorden: lijnen/kolommen van bolletjes ingekleurd, kruisen getrokken over antwoordschalen, ...
- Herhaaldelijk vermelden van opmerkingen die wijzen op een heel negatieve attitude ten aanzien van de vragenlijst. Bv. "da suckt de balle der echt af", "Da gadde u ni aan", ...
- Herhaaldelijk onmogelijke/verdachte antwoorden geven: "8 miljoen euro zakgeld per week", "zus heet kaka en woont in hoerenkot", ...

In totaal werden hierdoor 74 bijkomende datalijnen verwijderd.

Tenslotte worden de datalijnen gewist van respondenten die vermelden niet deel wensen te nemen aan het onderzoek (1 case).

Eindresultaat van deze cleaning is een dataset met 7035 datalijnen.

### **3. Datacleaning modules B en C (door eindgebruiker)**

De laatste fase van de datacleaning dient door de eindgebruiker zelf te gebeuren indien wordt gewerkt met gegevens uit module B en/of C van de vragenlijst(en). Door de volgorde en lengte van de vragenlijst ligt de algemene respons in modules B en C lager dan in module A. Naargelang de informatie waarmee wordt gewerkt kan door de eindgebruiker een additionele cleaning worden uitgevoerd voor respondenten die op respectievelijk module B of C een lage itemrespons hebben. Er werd hiervoor een variabele aangemaakt die de proportie item non-respons weergeeft op de basisvariabelen in deze modules (=variabelen aanwezig in de vier bevragsingsrondes<sup>2</sup>), dit zijn de variabelen propmissingsB en propmissingsC.

---

<sup>2</sup> Voor de verblijfskalender werd hiervoor enkel de variabelen met de informatie omtrent de eerste week gebruikt, omdat niet alle respondenten de volledige kalender dienden in te vullen.