SONO
STEUNPUNT
ONDERWIJS ONDERZOEK

DE INVLOED VAN SCHOOL- EN SYSTEEMKENMERKEN OP (ON)GELIJKE ONDERWIJSUITKOMSTEN NAAR SOCIALE HERKOMST EN THUISTAAL:

Vergelijkende analyse op PISA 2015

Franck, E., Ünver, Ö. Nicaise, I.

SCHOOL- AND SYSTEM-LEVEL DETERMINANTS OF SOCIOECONOMIC AND LANGUAGE-BASED INEQUALITIES IN COGNITIVE OUTCOMES

A comparative analysis based on PISA 2015

Franck, E., Ünver Ö. & Nicaise, I.

Promotor: Nicaise, I.

Research paper SONO/2019.OL1.3/2

Leuven: HIVA / Hamburg:IEA/ Gent: Vakgroep Onderwijskunde, 2019/11

KU LEUVEN HIVA

ONDERZOEKSINSTITUUT VOOR
ARBEID EN SAMENLEVING

GHENT
UNIVERSITY

IEA

Researching education, improving learning

ETN
OCCAM

www.etn-occam.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765400.

Het Steunpunt Onderwijsonderzoek is een samenwerkingsverband van UGent, KU Leuven, VUB, UA en ArteveldeHogeschool.

Gelieve naar deze publicatie te verwijzen als volgt:
Franck, E. & Nicaise, I. (2019) De invloed van school- en systeemkenmerken op (on)gelijke onderwijsuitkomsten naar sociale herkomst en thuistaal: vergelijkende analyse op PISA 2015. Leuven: HIVA / Hamburg: IEA/ Gent: Steunpunt Onderwijsonderzoek.

Voor meer informatie over deze publicatie Emilie.franck@kuleuven.be

Dit onderzoek werd deels gefinancierd door het Horizon 2020 programma van de EU, in het kader van het OCCAM netwerk (Marie Skłodowska-Curie project nummer 765400), en deels door de Vlaamse Gemeenschap, Ministerie voor Onderwijs en Vorming.

p.a. Coördinatie Steunpunt Onderwijsonderzoek
UGent – Vakgroep Onderwijskunde
Henri Dunantlaan 2, BE 9000 Gent

Deze publicatie is ook beschikbaar via www.steunpuntsono.be

Inhoud

Beleidssamenvatting

Gelijke onderwijskansen groeiden de voorbije decennia in vele landen uit tot een heikel thema binnen de onderwijsproblematiek. Het streven naar gelijke onderwijskansen houdt in dat schooluitkomsten van leerlingen niet (of zo weinig mogelijk) samenhangen met de sociaaleconomische status (SES), migratiestatus, thuistaal of gender van leerlingen. Vele onderwijssystemen hebben de voorbije decennia reeds heel wat middelen ingezet op het realiseren van gelijke onderwijskansen, waaronder Vlaanderen. Internationale studies zoals PISA en TIMSS tonen echter aan dat vele landen er tot op heden niet in slagen om de onderwijsongelijkheid tegen te gaan. Maar, er zijn wel grote verschillen tussen landen in de mate van onderwijsongelijkheid. Landen zoals o.a. Vlaanderen, de US of Hongarije staan bekend als onderwijssystemen met een grote onderwijsongelijkheid, terwijl landen zoals o.a. Denemarken, Finland en Canada het veel beter doen op vlak van gelijkheid. Niet alleen tussen landen worden verschillen geconstateerd, ook binnen een land blijken sommige scholen het beter te doen op vlak van gelijkheid dan andere. Dit alles leidt tot vragen zoals 'Hoe komt het dat sommige landen er beter in slagen de onderwijsgelijkheid te verbeteren dan andere?' of 'Waarom hebben sommige scholen een hogere mate van onderwijsongelijkheid dan andere?'.

Binnen het gelijke onderwijskansen debat, komt ook vaak de vrees naar voren dat meer gelijkheid ten koste gaat van de kwaliteit van een onderwijssysteem (Nicaise e.a., 2008). Uit een onderzoek van Danhier en Jacobs (2017) blijkt dit niet noodzakelijk het geval te zijn. Hun onderzoek toonde aan dat verschillende landen zoals Canada, Denemarken, Finland en Noorwegen simultaan beide doelen kunnen nastreven en realiseren. Deze landen combineren een hoge mate van sociale gelijkheid met hoge gemiddelde prestatieniveaus. Ondanks voorgaande voorbeelden zijn veel landen er nog niet in geslaagd de optimale verhouding tussen gelijkheid en kwaliteit te vinden en zitten zij in een trial-and-error proces. Ook in Vlaanderen is dit het geval.

Gegeven het feit dat vele onderwijssystemen zowel gelijkheid als kwaliteit hoog in het vaandel hebben, focussen we ons in dit rapport op beide zaken. Meer bepaald zullen we nagaan hoe school- en systeemkenmerken enerzijds de prestaties van leerlingen (hierna leerlingpresaties) en scholen (hierna schoolprestaties), en anderzijds de ongelijkheid tussen leerlingen (hierna leerling-ongelijkheid) en tussen scholen (hierna school-ongelijkheid) kunnen verklaren.

Algemeen hebben we vier onderzoeksvragen gesuggereerd:

(1) Hoe en in welke mate zijn de *leerlingprestaties* inzake wiskunde en leesvaardigheid van 15-jarigen geassocieerd met school- en systeemkenmerken?
(2) Hoe en in welke mate is de sociaaleconomische en linguïstische *ongelijkheid tussen leerlingen* geassocieerd met school- en systeemkenmerken?
(3) Hoe en in welke mate zijn de *schoolprestaties* inzake wiskunde en leesvaardigheid geassocieerd met school- en systeemkenmerken?
(4) Hoe en in welke is de sociaaleconomische en linguïstische *ongelijkheid tussen scholen* geassocieerd met systeemkenmerken?

Onze ongelijkheidsmaatstaven worden allen gemeten door middel van de SES-gradiënt en de taal-gradiënt. De sociaaleconomische en linguïstische leerling-ongelijkheid wordt gemeten door resp. (1) de correlatie (meer bepaald de beta-coëfficiënt) tussen de SES van leerlingen en hun wiskunde- en/of leesvaardigheidsprestaties, (2) de correlatie tussen de thuistaal van leerlingen (spreken leerlingen thuis dezelfde taal als op school?) en hun wiskunde- en/of leesvaardigheidsprestaties. De sociaaleconomische en linguïstische school-ongelijkheid wordt gemeten door resp. (1) de correlatie tussen de school SES en de schoolprestaties inzake wiskunde en leesvaardigheid (2) de correlatie tussen het percentage anderstaligen in een school en de schoolprestaties inzake wiskunde en leesvaardigheid.

In wat volgt bespreken we elk van onze onderzoeksvragen aan de hand van onze resultaten. Naast onze internationale schattingen, hebben we ook schattingen gemaakt voor Vlaanderen afzonderlijk (zie Annex 1). We stelden vast dat de resultaten voor Vlaanderen nauw aansluiten bij de resultaten die we constateerden op internationaal niveau.

Gemiddelde leerling- en schoolprestaties

In eerste instantie gingen we na welke school- en systeemkenmerken de leerling- en schoolprestaties inzake wiskunde en leesvaardigheid beïnvloeden.

Met betrekking tot de **schoolkenmerken**, constateren we allereerst dat *de kwaliteit van leerkrachten* positief gecorreleerd is met de leerling- en schoolprestaties inzake wiskunde en leesvaardigheid. Dit betekent dat leerlingen in scholen met meer gekwalificeerde leerkrachten, het beter doen dan leerlingen in scholen met minder gekwalificeerde leerkrachten. Het waarborgen van voldoende gekwalificeerde leerkrachten blijft met andere woorden een belangrijke taak van inrichtende machten en overheden.

Ten tweede vertoont ook de school-SES een positieve én sterke relatie met de leerling- en schoolprestaties. Dit suggereert dat de school SES een belangrijke determinant van de leerling- en school-ongelijkheid is: leerlingen in kansrijke scholen doen het aanzienlijk beter – ongeacht hun SES – dan leerlingen in kansarme scholen. Of nog, kansrijke leerlingen in een kansarme school zullen aanzienlijk lager scoren dan kansarme leerlingen in een kansrijke school. Het prestatieverschil tussen de meest kansrijke (95^e^ percentiel) en de meest kansarme (5^e^ percentiel) scholen in de OESO wordt in onze analyses geschat op circa 3 jaar formeel onderwijs. Alhoewel de 'invloed' van de school-SES niet te miskennen is, willen we toch voorzichtig zijn wat de richting van het effect betreft: enerzijds beïnvloedt de klassamenstelling de prestaties van individuele leerlingen; maar anderzijds bepalen hun prestaties ook de klas waar ze in terechtkomen. Sociale herkomst, schoolse segregatie en cognitieve prestaties beïnvloeden elkaar wellicht in een soort van escalerende beweging, waarbij ze doorheen de schoolloopbaan cumulatief gaan werken. Dit betekent meteen dat de variabele 'school SES' niet het momentaan effect van de huidige school- en klassamenstelling meet, maar eigenlijk als proxy dient voor de vicieuze cirkel waarin leerlingen uit lagere socioeconomische milieus van bij de start van hun loopbaan terechtkomen. We onderscheiden in de literatuur 7 mogelijke mechanismen die deze positieve correlatie tussen school SES en prestaties kunnen verklaren:

1) Peer group effecten: leerlingen leren niet alleen van de leerkrachten maar ook van elkaar. Het concentreren van kansarme leerlingen (die doorgaans zwakker presteren) in eenzelfde school

is daarom nadelig aangezien ze minder van elkaar zullen leren (Thrupp et al., 2002; Van Eewijck & Slegers, 2010).
2) De 'demotie-ervaring' die leerlingen in kansarme scholen meemaken werkt eveneens 'demotivatie' in de hand (Pelleriaux, 2001; Van Houtte & Stevens, 2010; Spruyt, 2014).
3) Leerkrachten hebben de neiging om tegemoet te komen aan de zwakkere prestaties van hun leerlingen door de lat lager te leggen en minder inspanningen te vergen (Rosenthal & Jacobson, 1968; Heyerick, 1978; Speybroeck, 2013).
4) Schoolklimaat-effecten: concentratiescholen hebben in de regel meer tuchtproblemen, leerlingen zijn meer gefrustreerd, leerkrachten passen hun verwachtingen neerwaarts aan en spenderen meer tijd aan klasmanagement, ten koste van effectieve instructietijd (Thapa et al., 2013; Gustafsson et al., 2018).
5) Ongelijkheden in het sociaal en cultureel kapitaal van de schoolomgeving: concentratiescholen kunnen minder beroep doen op de sociale en culturele hulpbronnen van ouders en van de lokale gemeenschap (die zelf vaak kansarm is – Poesen-Vandeputte & Nicaise, 2015).
6) Soortgelijke Matteüseffecten spelen in het menselijk kapitaal en de materiële middelen van scholen: kansarme scholen worden vaak geconfronteerd met meer verouderde infrastructuur en minder sterke en ervaren schoolteams (Thrupp & Lupton, 2006; OECD, 2018b; Poesen-Vandeputte & Nicaise, 2015). Bovendien hebben kansarme scholen in gedecentraliseerde onderwijssystemen ook vaak minder werkingsmiddelen ter beschikking, omdat de lokale overheden die ze financieren zelf over kleinere budgetten beschikken.
7) Verschillen in curriculum tussen scholen met een hoge versus lage gemiddelde SES van leerlingen: Hier denken we vooral aan de 'tracking' die zo typisch is voor het secundair onderwijs. Leerlingen uit lagere sociale lagen van de bevolking worden systematisch meer naar het beroepsonderwijs georiënteerd, terwijl leerlingen uit begoede milieus in het algemeen vormend onderwijs geconcentreerd zitten. Het algemeen vormend onderwijs bevat uiteraard veel méér vorming in sleutelvaardigheden zoals wiskunde en taal dan het beroepsgericht onderwijs. Ongeacht de capaciteiten van de leerlingen zal dit verschil in aanbod de ongelijkheid in vaardigheden versterken (OECD, 2016a; Lavrijsen & Nicaise, 2015; Dockx, 2019). Analoge mechanismen spelen in systemen met een comprehensief curriculum waar de leerlingen op basis van hun prestaties gesegregeerd worden in niveauklassen.

Alles samengenomen, is de school-SES (als uitkomst van een geleidelijk segregatieproces doorheen de voorbije schoolloopbaan) sterk gecorreleerd met de leerling- en schoolprestaties. Deze vaststelling riskeert uiteraard – door de quasimarktwerking in ons onderwijs – de segregatie en de ongelijkheid verder in de hand te werken: kansrijke ouders zullen immers vermijden om hun kinderen naar minder bevoorrechte scholen te sturen, tenzij de sociale mix door de overheid gereguleerd wordt.

Anderzijds vertoont het *percentage anderstaligen in een school* een negatieve, maar zeer zwakke relatie met de (gemiddelde) wiskunde- en leesvaardigheidsprestaties van 15-jarige leerlingen. We concluderen aldus dat het percentage anderstalige leerlingen op school – na controle voor de individuele thuistaal - geen substantiële invloed heeft op de leerling- en schoolprestaties.

Tot slot, is de *leerling-leraar ratio* positief – maar zeer zwak – gecorreleerd met de schoolprestaties, en negatief – maar zeer zwak – met de leerlingprestaties. Gezien de verwaarloosbaar kleine correlaties, besluiten we dat de leerling-leerkracht ratio niet geassocieerd is met leerling- en schoolprestaties.

Met betrekking tot de **systeemkenmerken**, zien we dat twee systeemkenmerken positief gecorreleerd zijn met de leerling- en/of schoolprestaties: *BBP per capita* (= inkomen per hoofd van de bevolking) en de *competitiviteit tussen scholen*. Dit betekent dat leerlingen in landen met een hogere economische welvaart, en landen met meer competitie tussen scholen, beter scoren dan leerlingen in landen met minder economische welvaart en een lager niveau van competitiviteit tussen scholen.

De leeftijd waarop *leerlingen worden georiënteerd ('getracked')* vertoont een negatieve (maar nauwelijks significante) relatie met de leerlingprestaties in wiskunde. Voor leesvaardigheid is de relatie niet significant; en hetzelfde geldt voor de relatie met de gemiddelde schoolprestaties inzake wiskunde en leesvaardigheid. Deze bevindingen willen we echter relativeren om verschillende redenen: onder de landen met 'late tracking' zijn er een aantal minder ontwikkelde (niet-OESO) landen waar helemaal geen tracking plaatsvindt en waar het onderwijs in het algemeen zwakker is, wat mogelijks de negatieve coëfficiënt voor wiskunde verklaart (Lavrijsen, 2013). Bovendien is deze laatste slechts significant op 10%-niveau. Een correcter beeld wordt bekomen als gecontroleerd wordt op leerlingprestaties aan het eind van het lager onderwijs. In Lavrijsen & Nicaise (2015; 2016) werd gecorrigeerd voor de verdeling van prestaties in het vierde leerjaar van het lager onderwijs, en bleek het effect van latere tracking op de gemiddelde vaardigheden op 15 jaar eerder *positief* te zijn, zowel voor leesvaardigheid als voor wiskunde en wetenschappen. De enige uitzondering hierop geldt voor de 25% top-presteerders in wiskunde, die meer baat schijnen te hebben bij vroege tracking – maar deze baat is niet statistisch significant.

Tot slot, tonen de andere systeemkenmerken – *de frequentie van zittenblijven, de overheidsuitgaven aan basisonderwijs* – geen verband met de gemiddelde leerling- en schoolprestaties.

Sociaaleconomische en linguïstische leerling-ongelijkheid in prestaties

Hoe beïnvloeden de school- en systeemkenmerken de ongelijkheid in prestaties tussen leerlingen, naargelang hun socioeconomische achtergrond en thuistaal? We beperken ons tot het bespreken van de belangrijkste resultaten. Opnieuw focussen we eerst op de schoolkenmerken en nadien op de systeemkenmerken.

Met betrekking tot de **schoolkenmerken**, observeren we in eerste instantie dat twee schoolkenmerken – *de leerling/leraar ratio en het percentage anderstaligen op school* – niet of verwaarloosbaar zwak geassocieerd zijn met de sociaaleconomische en linguïstische leerling-ongelijkheid. Dat laatste is verrassend, omdat men a priori zou verwachten dat de taalkloof moeilijker te overbruggen is in scholen waar het aandeel anderstaligen hoger is (en de interactie met 'lokale' leerlingen dus beperkter is). Toch komen onze bevindingen overeen met ander onderzoek op dit domein. Een eerste mogelijke verklaring is het zogenaamde 'specialisatie-argument', nl. dat scholen met veel anderstaligen hun taalonderricht ook op deze doelgroep kunnen afstemmen, en meer resultaat bereiken dan meer gemengde scholen. Daarnaast moet men echter ook rekening houden met het feit dat het interactie-effect tussen 'aandeel anderstaligen' en de taalkloof geldt na uitzuivering van SES-effecten en het individuele hoofdeffect van anderstaligheid. Toch is dit een opmerkelijke vaststelling, die de beeldvorming over concentratiescholen als 'scholen met veel allochtone leerlingen' nuanceert. Het feit dat scholen met een hoge concentratie van leerlingen met een migratie-achtergrond minder effectief zijn zou eerder te maken hebben met de lage school SES dan met het grote aandeel anderstaligen.

Ten tweede zien we dat een hoger *aandeel gekwalificeerde leerkrachten* de linguïstische leerling-ongelijkheid inzake wiskunde verkleint. De sociaaleconomische leerling- en school-ongelijkheid is daarentegen niet gerelateerd met de kwaliteit van leerkrachten. Desalniettemin kunnen we op basis van het eerstgenoemde effect (kleinere taalkloof) stellen dat het waarborgen van gekwalificeerde leerkrachten een effectieve investering is: meer gekwalificeerde leerkrachten verhogen niet alleen leerling- en schoolprestaties, maar verkleinen ook de linguïstische leerling-ongelijkheid.

Tot slot, is er een positieve interactie tussen de (cumulatieve) *school SES* en de sociaaleconomische leerling-ongelijkheid. Dit betekent dat kansrijke scholen méér sociaaleconomische leerling-ongelijkheid hebben dan kansarme scholen, of nog dat het prestatieverschil tussen kansarme en kansrijke leerlingen in kansrijke scholen groter is dan het prestatieverschil in kansarme scholen. Alhoewel het tegenovergestelde resultaat vaak (impliciet) verwacht wordt, vertonen onze resultaten en enkele voorgaande studies een positieve coëfficiënt. We denken aan twee mogelijke mechanismen die deze coëfficiënt kunnen verklaren:

1) Stigmatisering: kansarme leerlingen worden mogelijks sterker gestigmatiseerd of gediscrimineerd in kansrijke scholen. Dit zou dan leiden tot een lager schoolwelbevinden, een lagere motivatie en een minder goed zelfbeeld, wat de leerlingprestaties van kansarme leerlingen benadeelt.
2) Het sociaal en cultureel kapitaal van leerlingen: leerkrachten hebben in kansrijke scholen vaak hogere verwachtingen en gaan ervan uit dat de leerlingen de leerstof snel oppikken. Kansrijke leerlingen kunnen – bij moeilijkheden – gebruik maken van het sociaal en cultureel kapitaal van hun familie en/of lokale gemeenschap. Dit is vaak niet het geval voor kansarme leerlingen.

Wat de verklaring dan wel mag zijn, het feit dat we een positieve associatie constateren suggereert dat schoolsegregatie de sociaaleconomiche leerling-ongelijkheid versterkt. Hieruit kunnen we besluiten dat het inzetten op een betere sociale mix in scholen en dus het tegengaan van schoolsegregatie door middel van directe (bv. Inschrijvingsbeleid) en indirecte (bv. leerlingen later tracken) maatregelen, een streefdoel van overheden moet zijn. Het positieve interactie-effect geldt echter niet voor anderstalige leerlingen: de samenhang tussen de school SES en de linguïstische leerling-ongelijkheid is ambivalent.

Van alle **systeemkenmerken** in onze analyses, vertoont enkel *de leeftijd waarop leerlingen getracked* worden een significante correlatie met de ongelijkheid in cognitieve prestaties. Meer bepaald vinden we een negatieve relatie met de linguïstische en sociaaleconomische leerling-ongelijkheid. Dit betekent dat in onderwijssystemen waar leerlingen op een latere leeftijd getracked worden, zowel de sociaaleconomische als linguïstische leerling-ongelijkheid kleiner is. In voorgaande studies wordt dit vaak verklaard door het feit dat oudere leerlingen minder beïnvloed worden door hun ouders en peers bij het maken van een studiekeuze. Op jongere leeftijd kiezen ouders vaak de studierichting, en deze keuze is meestal sterk beïnvloed door hun eigen achtergrond. Hierdoor zullen kansarme ouders hun kinderen sneller naar 'hiërarchisch' lagere tracks sturen dan kansrijke ouders. Bovendien krijgen zwakkere leerlingen bij latere tracking meer algemene vorming, waardoor ze op 15-jarige leeftijd beter presteren op wiskunde en taal. Vanuit dit opzicht is het later tracken van leerlingen een effectieve manier om de leerling-ongelijkheid tegen te gaan, alsook de schoolsegregatie.

Alle andere systeemkenmerken – de frequentie van zittenblijven, BBP per capita, overheidsuitgaven aan lager onderwijs en de competitiviteit tussen scholen – tonen geen of slechts een heel zwakke correlatie met de ongelijkheid op leerlingniveau.

Sociaaleconomische en linguïstische school-ongelijkheid

In een laatste deel zijn we ingegaan op de vraag welke systeem-kenmerken samenhangen met de ongelijkheid tussen *scholen*.

Onze resultaten wijzen (enigszins verrassend) op geen enkel systeemkenmerk dat substantieel gerelateerd is aan de school-ongelijkheid. Hier en daar zagen we een verwaarloosbaar kleine correlatie. De zwakke verklaringskracht van de geselecteerde variabelen zou eventueel kunnen toegeschreven worden aan het gebruik van te ruwe maatstaven. Bovendien kunnen andere factoren die niet in ons model zijn opgenomen (zoals bv. decentralisatie of verschillen in bestuurlijke capaciteit van scholen) mogelijk wel relevant zijn. Verder onderzoek op dit vlak blijft dus aanbevolen.

Implicaties

Onze bevindingen hebben heel wat mogelijke beleidsimplicaties. Omdat dit rapport kadert in een onderzoek naar de effectiviteit van het GOK-ondersteuningsbeleid bespreken we in deze sectie eerst algemene implicaties m.b.t. gelijke onderwijskansen, en vervolgens implicaties m.b.t. positieve discriminatie in de schoolfinanciering (de zogenaamde *equity funding*).

Implicaties m.b.t. het gelijke onderwijskansenbeleid in het algemeen

Allereerst blijkt dat het garanderen van voldoende gekwalificeerde leerkrachten een cruciale investering is voor overheden, wil men de ongelijkheid tegengaan en de prestaties verbeteren. Temeer gezien het risico bestaat dat tekorten op de lerarenmarkt Mattheüs-effecten teweegbrengen: vaak zijn kansarme scholen de eerste slachtoffers van een tekort aan gekwalificeerde leerkrachten, omdat ze minder aantrekkelijk zijn voor leerkrachten in het algemeen.

Ten tweede, doen onderwijssystemen waar leerlingen op latere leeftijd getracked worden het duidelijk beter inzake onderwijsgelijkheid. Bovendien is de schoolsegregatie ook vaak kleiner in deze landen. Een overgang naar een meer comprehensief onderwijsssyteem – in landen waar leerlingen op vroege leeftijd getracked worden – is met andere woorden aan te raden om de onderwijsongelijkheid tegen te gaan. Het tegenargument, nl. de vrees dat zo'n hervorming een neerwaartse nivellering zou teweegbrengen, wordt door onze schattingen schijnbaar deels bevestigd: ceteris paribus liggen de gemiddelde prestaties van leerlingen in landen met een latere tracking niet significant lager voor taal, maar wel voor wiskunde. De beta-coëfficiënt voor wiskunde is echter slechts 'marginaal significant'. Zoals we hogerop reeds aangaven, wordt de hypothese van neerwaartse nivellering bij latere tracking in vroegere, meer diepgaande studies grotendeels verworpen, en is er zelfs sprake van *opwaartse* nivellering. De enige uitzondering hierop geldt voor de 25% top-presteerders in wiskunde. Voor hen zou in systemen met latere tracking een intensiever aanbod wiskunde kunnen aangeboden worden, om het 'nadeel' van latere tracking te voorkomen.

Ten derde, wijst de sterke samenhang tussen school-SES en leerling- en schoolprestaties, op de noodzaak van meer sociaal gemengde scholen. Niet alleen hangen de individuele leerlingprestaties méér samen met de school-SES dan met de individuele SES (additief effect); bovendien stellen we ook vast dat binnen hoge SES scholen de sociaaleconomische ongelijkheid in cognitieve prestaties uitvergroot wordt (multiplicatief effect). Wederom moeten we concluderen dat het nastreven van sociaal gemixte scholen van groot belang is wil men de ongelijkheid tegengaan. De instrumenten hiervoor zijn bekend: het inschrijvingsdecreet, het terugdringen van het zittenblijven, het inclusief onderwijs, de afschaffing van de schotten tussen onderwijsvormen in het secundair onderwijs, de versterking van het GOK-ondersteuningsbeleid, enzovoort.

Implicaties m.b.t. het GOK-ondersteuningsbeleid in het bijzonder

Het doel van onze analyses was om de discussie inzake de effectiviteit van het gelijkekansenbeleid – en meer bepaald het GOK-ondersteuningsbeleid in Vlaanderen – te kaderen in een ruimer, systemisch kader. We zagen immers dat verschillen in school- en systeemkenmerken kunnen verklaren waarom het gelijkekansenbeleid in sommige landen effectiever is dan in andere: de effectiviteit van positieve discrimatiemaatregelen hangt namelijk af van het relatieve gewicht van dergelijke maatregelen in vergelijking met andere systeemkenmerken, die de kansengelijkheid kunnen bevorderen of belemmeren.

De belangrijkste boodschap van onze analyses is het sterke verband tussen de school SES en de cognitieve prestaties van leerlingen. Het effect van de individuele SES van leerlingen kan immers worden "versterkt" door segregatiemechanismen die ertoe leiden dat leerlingen in verschillende scholen worden gesorteerd op basis van hun SES. Dit zal uiteindelijk leiden tot meer ongelijkheid in schoolprestaties. Bovendien vinden we een positieve interactie tussen de individuele SES en de school-SES, wat betekent dat er meer ongelijkheid is in elitescholen (hoge SES scholen) dan in gemiddelde of lage SES scholen. Dit heeft drie directe implicaties voor het GOK-ondersteuningsbeleid: (a) sociale segregatie vergroot ongelijkheden in de schoolprestaties en daarmee de *behoefte* aan bijkomende financiële middelen voor kansarme scholen, (b) het "versterkende effect" van schoolsegregatie verklaart mogelijks de lagere *doeltreffendheid* van positieve discrimatiemaatregelen in een meer gesegregeerd schoollandschap, (c) indien positieve discrimatiemaatregelen inzake schoolfinanciering gebruikt worden om de prestatiekloof tussen gesegregeerde scholen te dichten, lijkt het rechtvaardig om de financiering per kansarme leerling te 'versterken' met een coëfficiënt die omgekeerd evenredig is de *gemiddelde school-SES*. Laat ons dit illustreren met een concreet voorbeeld. Stel dat er twee scholen zijn met een even groot absoluut aantal leerlingen uit kansengroepen; en stel dat deze leerlingen 80% van de schoolpopulatie vertegenwoordigen in school A, en 50% van de leerlingpopulatie in school B; dan zouden we uit onze bevindingen concluderen dat het (ceteris paribus) verantwoord is om meer bijkomende middelen toe te wijzen aan school A. In de huidige regeling krijgen beide scholen echter evenveel 'extra'.

Het argument hierboven betekent ook dat – bedoeld of onbedoeld – financiële compensatiemechanismen vaak een eerder *remediërende* rol spelen aangezien ze gericht zijn op het 'wegwerken' van ongelijkheden die door andere mechanismen worden gegenereerd. Een meer preventieve aanpak kan dus voordeliger zijn: het *voorkomen* van segregatie op school (bijvoorbeeld door de regulering van de verdeling van leerlingen over scholen) kan de budgettaire kost van het gelijkekansenbeleid substantieel verlagen. Toegegeven, in landen (zoals België) waar onderwijsvrijheid en vrije schoolkeuze als grondwettelijke rechten worden beschouwd, stuiten zulke maatregelen op veel weerstand. Desalniettemin is belangrijk om te beseffen dat deze vrijheid een prijskaartje heeft.

Tot slot, suggereert recent onderzoek (OECD, 2014; 2018b) dat de 'kwaliteit' van leerkrachten (gemeten aan de hand van hun kwalificatieniveau) belangrijker is voor het verbeteren van de prestaties van (kansarme) leerlingen dan de 'kwantiteit' van leerkrachten (gemeten aan de hand van de leerling/leraar ratio). Onze bevindingen sluiten hierbij aan: het effect van de leerling/leraar ratio lijkt op schoolniveau geen wezenlijke invloed te hebben op de resultaten van de leerlingen - noch gemiddeld, noch op de verdeling ervan.[1] Dit betekent concreet dat het GOK-ondersteuningsbeleid zich eerder zou moeten richten op het aantrekken – en de verankering - van beter gekwalificeerde en meer ervaren leerkrachten dan op het uitbreiden van lestijdenpakketten.

[1] De gemiddelde leerling/leraar ratio op schoolniveau is weliswaar een vrij ruwe maatstaf. Zoals hogerop reeds werd aangestipt, is een meer gedetailleerde analyse wenselijk om de optimale kwantiteit-kwaliteitsmix te bepalen, zodat het rendement van de extra-investering in personeel gemaximaliseerd wordt.

Introduction

During the past few decades, the principle of 'equity in education' has gained much attention and became a fundamental value and guiding principle in education policy (OECD, 2018). Since the work of James Coleman (1968) on equal educational opportunities, ample evidence showed that children's educational success is not only determined by their individual ability and effort, but by a multitude of factors outside of their control such as their socioeconomic status (SES), ethnicity, gender, the school they are attending, social or economic circumstances of the country, etc. (Pfeffer, 2008; Betts and Roemer, 2005; Li, 2010). This was and is considered as problematic by many policy-makers, government officials and scholars. Education is after all believed to be a crucial lever for one's future income and productivity, and thus, society's development and growth. Hence, it is seen as one of the most effective tools in facilitating upward social mobility (Roemer, 1998; OECD, 2018).

As we live in times with high economic inequalities, improving equity in education has even become a more urgent matter (OECD, 2018). Income inequality in developed countries is at its highest since the 1980s (OECD, 2015d; OECD, 2011a). Research on social mobility shows that greater income inequality limits educational opportunities for talented but disadvantaged students (Lee and Lee, 2018). Consequently, high levels of income inequality go hand in hand with high levels of educational inequalities, which eventually will not only be detrimental – especially – for disadvantaged children, but for the society in general (Woessmann, 2009; Leuven et al., 2007).

Given the democratic principle that everyone deserves a fair and equal chance to improving their lives, many societies have designed policy programmes to tackle educational inequalities, namely 'educational priority policies' (EPPs). These aim to compensate for the educational disadvantage of less privileged populations, and mark society's acknowledgement that students should not be impeded by circumstances outside of their control. EPPs focus on supporting schools and/or education systems in their effort to offer more opportunities to children and students born in disadvantaged families (OECD, 2018).

Within the educational field, however, it is feared that improving educational equity goes at the expense of average performance. This is part of a wider debate on the so called 'efficiency-equity trade-off'. Often, politicians believe that they need to choose between boosting efficiency – and hence boosting average performance of students with a minimum amount of resources – or boosting equity by prioritising disadvantaged groups. Several studies suggest that this is a false dilemma (Woesmann et al., 2006): countries can promote efficiency and equity simultaneously. In fact, more equitable education systems tend to perform (on average) better than non-equitable educational systems (e.g., OECD, 2013a).

Unfortunately, despite the efforts of many countries to improve equity, international studies show that no country in the world can yet claim to have eliminated all social and language-based inequalities in education (OECD, 2018). There are however countries that do succeed better than others in increasing equity in education, and within countries, some schools have higher levels of equity than others (OECD 2018; Cumberworth et al., 2018; Gustafsson et al., 2018). These differences between countries and schools prove that there is nothing inevitable about disadvantaged students performing worse than their more advantaged counterparts (OECD, 2018). So far, little attention has been devoted to investigating which school- and system-level characteristics possibly influence the socioeconomic and language-based equity. This knowledge is required if one aims to reduce educational inequity.

In this report, we contribute to the literature by examining how school- and system-level characteristics are correlated with students' and schools' academic performance (educational success), as well as with educational inequity amongst students (student-inequity) and schools (school-inequity). We hope to gain a better understanding on why some schools and/or education systems succeed better than others in improving their educational success and educational inequity. This information can be valuable for policy-makers as it can be used as an indication of the extent to which education policies (for instance EPPs) should target and invest in specific school- and system-level characteristic in their effort of eliminating educational inequalities and boosting the performance of students.

In what follows, we start by elaborating on the concept of educational equity (§1). In the second section (§2) we will discuss – based on an extended literature review – the most important student –, school- and system-level characteristics that are correlated with the educational success and the degree of educational inequity of students and schools. In this section, our measures of educational inequity will be discussed more in depth. Next (§3), our research questions will be presented, followed by the data and method. In the fourth section (§4) the results of the analyses will be discussed. Finally, a discussion and conclusion will be provided in the last section (§5).

1. Equity in Education

Within the education literature, equity is not a simple and straightforward concept (Perry, 2007; 2009). It is used and measured in various – and sometimes conflicting – ways in research, policies and social and public debates. Providing an extended overview of all the conceptualizations of 'equity' and its historical roots, falls beyond the scope of this article.

In this article, equity entails that educational outcomes should be unrelated to circumstances beyond a person's control (such as socioeconomic status, ethnicity, gender, the school they are attending, social or economic circumstances of the country, etc.) (OECD, 2018; Perry, 2009). In other words, while differences in educational outcomes based on the individual abilities of students are accepted, inequalities resulting from the life circumstances of the students - exogenous as well as endogenous to the education system - should be minimized.

2. Educational success and equity: which characteristics do matter?

As mentioned above, the aim of this report is to understand which school- and system-level factors are associated with the educational success (measured by students' - and schools' average - performance in mathematics and reading performance) and the degree of educational inequity amongst students and schools. The reasoning behind this is that some system-level characteristics might only be associated with student-inequity (the degree of inequity on student level) or school-inequity (the degree of inequity on school-level). For instance, it could be the case that a country's percentage of students that have repeated a grade is only correlated with student-inequity, but not with school-inequity.

In general, the focus of this report is threefold; (1) Which school- and system-level factors are correlated with students' and schools' educational success and to what degree?; (2) how does student-inequity vary with school- and system-level characteristics?; (3) how does school-inequity vary with system-level characteristics?

Two parameters are used to measure student-inequity: (1) the SES-gradient which refers to the correlation between students' SES and their performance, and (2) the language-gradient which refers to the correlation between students' language spoken at home (whether or not it is the same as the one spoken at school) and their performance. By including these two measures, we capture both the socioeconomic and the language-based inequity amongst students. Similarly, school-inequity is estimated by two measures: (2) the school SES-gradient which refers to the correlation between a schools' average SES and the average performance of students in that school, and (2) the language-gradient which refers to the correlation between the percentage of students that speak another language at home than the instruction language at school and the average performance of students in a school. Again, with these measures we capture both socioeconomic and language-based inequity amongst schools.

2.1. Student-level characteristics

Students' educational success is in all countries – although to different degrees – not only determined by their individual ability, but by a multitude of factors outside of their control (Pfeffer, 2008; Betts and Roemer, 2005; Li, 2010; OECD, 2018). The most important factors on student-level are: socioeconomic status, immigration status, language spoken at home, gender, age, grade repetition and program orientation[2].

[2] Grade repetition and program orientation will not be included and discussed in this study due to their endogenous character: they do not only influence students' performance and student-equity, but are also determined by them.

2.1.1. The socioeconomic status

Socioeconomic status can be defined as “a person’s overall social position… to which attainments in both the social and economic domain contribute” (Ainley et al., 1995: ix). When used in studies of student performance, it refers to the SES of the students’ parents. Different studies used different measures of SES, but typically, it is a combined measure based on the income of parents, their occupation, their educational attainment and sometimes the home resources. The income of parents is a first indicator of the SES and reflects the potential for social and economic resources that parents have and that are available to the student. Second, the parental education reflects the social and cultural capital of parents and is also an indicator of parent’s income. Third, the occupation of parents represent information not only about the income and education required for an occupation but also about the prestige and culture of a given socioeconomic stratum. Fourth, the home resources, is not used that commonly as the other three main indicators. In recent years, however, researchers have emphasized the significance of this indicator. Home resources include household possessions such as books, computers, and a study room, as well as the availability of educational services after school and in the summer (Siring, 2005).

A large body of empirical evidence has established student family SES as being one of the most powerful student-level characteristic that is positively correlated with student performance and educational inequity (Gustafsson et al., 2018; Cumberworth et al., 2018). More specifically, the stronger the correlation between students’ SES and their cognitive outcomes, the higher the socioeconomic inequality in a school or education system. The strength of the gradient does vary considerably across countries and schools, meaning that some countries and schools succeed better in achieving socioeconomic equity than others. Yet, International large scale assessments revealed that there is not a single country that can yet claim to have eliminated the correlation between students’ SES and their performance (OECD, 2016a; OECD, 2018). A PISA report demonstrated for instance that in 2015, the SES of 15-year-old students explains – on average across all OECD countries – 12.9% of the variation in science performance within each country. In some countries the strength of this SES-gradient is higher and accounts for more than 20% of the variation in students’ performance (for instance in Hungary, Luxembourg and Peru), whereas in other countries it explains less than 10 % of the variation (for example in Canada, Iceland, Latvia, Norway, Turkey, etc.). (OECD, 2016a; OECD, 2018; Vakgroep Onderwijskunde, 2015).

Similar conclusions can be drawn for primary education. Studies using TIMSS[3] data show for example that there is a positive correlation between the SES of students in the fourth grade and their science and mathematics scores. Again, the correlation varies considerably between countries and schools, meaning that in some countries and schools students’ SES explain more (less) variation in students’ performance (Bellens 2013a, 2013b; Vandenbroek et al., 2016; Caperona et al., 2016).

In this report, (the strength of) the SES-gradient will be used to measure the socioeconomic inequity amongst students.

2.1.2. Language spoken at home

A second important student-level characteristic that is strongly correlated with student performance and will be used as a measure of language-based student-inequity, is the mother tongue of a student.

[3] Trends in International Mathematics and Science study

Speaking a different language at home than the one used at school is one of the barriers to learning that students with (often) immigrant background and other students must try to overcome (OECD, 2016b; OECD 2013a). Moreover, linguistic minority students may have less friends and therefore also less social capital as a source of support in school-related matters. PISA results show for example that in 2012 – on average across OECD countries – the odds of low performance in mathematics among students who speak a different language at home are more than twice than high as the odds among students who do speak the same language. In other words, being a non-native student is negatively correlate with students' cognitive outcomes. The stronger this correlation, the larger the achievement gap between native and non-native students and hence the larger the language-based inequity amongst students.

Despite evidence on a rather strong correlation in many countries, research also found that the strength of the correlation reduces after controlling for students' SES and migration status. However, in some countries (such as the Flemish Community of Belgium), the effect of language remains strong, even after controlling for other student-level characteristics such as SES and migration status (OECD, 2016b; OECD, 2013a; Frank & Nicaise, 2019).

2.1.3. Migration status

Students with an immigrant background differ by country of origin, culture and mother tongue, and bring a wide range of skills, knowledge and motivations to their schools. Even though a significant number of migrants are highly skilled, most of them are socio-economically disadvantaged (OECD, 2013a).

Many studies have shown that non-immigrant students tend to perform immigrant students in most countries (in a few countries such as Australia, New Zealand and Singapore, the opposite is observed), with first-generation immigrant students performing worse than second-generation immigrant students. Immigrant students also more frequently attend schools with fewer resources and higher concentrations of disadvantaged students, are more likely than their non-immigrant peers to repeat a grade and are more often enrolled in vocational tracks (OECD, 2015a; 2016a; OECD, 2016b). Hence, there is a negative relationship between being an immigrant student and cognitive outcomes.

The influence of migration status on cognitive outcomes is closely intertwined with the effect of the language spoken at home and the SES of students. Hence, in many countries the correlation between students' migration status and their academic performance reduces considerably or becomes non-significant after controlling for SES and language spoken at home (OECD, 2016b; OECD, 2013A). Although the correlation with migration status remains significant in some countries, indicating that the SES and language spoken at home cannot fully account for immigrant students' poorer performance (Vakgroep Onderwijskunde, 2015; OECD, 2016A), we decided to use the language spoken at home as our measure for language-based student-inequity as its correlation with student performance is often stronger.

2.1.4. Gender

It has been repeatedly noted that girls outperform boys in reading, whereas boys outperform girls in mathematics and science (OECD, 2016b, 2013a; OECD, 2016a; Jacobs et al., 2007). These gender differences in achievement are not solely explained by innate ability, but by social and cultural context factors that reinforce stereotypical attitudes and behaviours (OECD, 2015c). For example, boys are more likely than girls to be disengaged from school and to play video games in their free time, whereas girls spend more time doing homework, tend to behave better in class and read more complex texts (like fiction) in their free time. On the other hand, girls have lower levels of self-efficacy in mathematics and science (OECD, 2015c; OECD, 2016b; OECD, 2013a).

2.1.5. Age

In most education systems, the difference in age between the youngest and oldest pupil in a classroom is – without taking into account students that have repeated a grade – maximum one year. It is assumed that children within a year group have corresponding levels of maturation and experience, and can be educated together.

Many studies that have focused on the effect of the date/month of birth on cognitive test scores found a significant positive effect of age on school performance, with older children outperforming their younger peers (Bedard et al., 2006; Solli, 2017; Crawford et al., 2007). Other studies also conclude that younger students have a higher risk of repeating a grade, are more often referred to special education and have more academic or behavioral problems (Wallingford et al., 2000; Wilson, 2000; Tarnowski et al., 1990; Verachtert et al., 2008).

2.2. School-level characteristics

School-level characteristics can have a considerable influence on students' performance and the level of educational inequity amongst students. In the following section, we will discuss the most important ones.

2.2.1. The socioeconomic status of the school

Many studies have investigated the relationship between schools' socioeconomic profile (a proxy for socioeconomic school composition and school segregation) and student performance. Within the literature, there is much disagreement about the operationalization of a school composition effect and the reliability of its effect-size (Dicke et al., 2018; Van Erwijk et al., 2010; Cumberworth et al., 2018). First, there is an issue of measurement error in the individual SES, whereby average school SES takes over part of the individual effect. Admittedly, little can be done about it, apart from mentioning that individual SES effects may be under-estimated and school SES over-estimated. A second discussion relates to so called "phantom-effects". More specifically, it is argued that the 'pure' school composition effect can only be correctly measured if the model accounts for unequal 'starting positions', i.e. when students access their school in the first grade. Without this correction, it is argued, school SES will capture the influence of antecedents in the school career that are just correlated with SES. In our view, this 'phantom effect' argument depends on the purpose of each research. In the pedagogical literature, where research mainly revolves around the determinants of learning progress within a clearly delimited period of time (e.g. one school year), the effects of pre-existing inequalities in cognitive skills should obviously be filtered out so as to measure the net effect of individual and school characteristics during the observed period. However, our research focuses on systemic effects throughout the school career: it aims to capture the *cumulative* impact of social inequalities on cognitive skills at age 15, whatever their origin (social capital acquired in the students' home environment as well as segregation during the entire past school career) and whatever the stage at which they were shaped. School SES at age 15 should then be interpreted as a proxy for the students' shared social capital, which is (at least partly) determined by progressive segregation mechanisms rooted in the education system.

Despite disagreement within the educational field about the measurement and interpretation of the socioeconomic school composition effects (Marks, 2015), the bulk of the evidence continues to suggest that it matters more than one is willing to accept (Van Erwijk et al., 2010, Cumberworth et al., 2018). With respect to the correlation between a school's socioeconomic profile and students' performance, most studies observe a positive relation in both primary and secondary education, meaning that a student attending a school with a high SES profile will perform better than that same student attending a school with a low SES profile (Agirdag et al., 2012; OECD 2018). A recent PISA report (2018) shows for example that – on average across all OECD countries – disadvantaged students who attend advantaged SES schools score 78 points higher on science than their disadvantaged peers who attend disadvantaged SES schools (OECD, 2018). This is the equivalent of nearly two years of formal schooling[4]. The relation between the SES profile of a school and student performance is however not equal for all students. Disadvantaged students attending disadvantaged schools are, a priori, disadvantaged twice as they strive for achievement. Apart from the gap in learning opportunities and performance they already face due to their SES and mother tongue, disadvantaged schools are more often confronted with difficult learning environments, lower teacher expectations, lower levels of competition between the students, the lack of resources, weaker parental support, the lack of qualified teachers and management staff, etc. These school-characteristics also have a significant negative impact on disadvantaged students' performance (OECD, 2016A; OECD, 2018; OECD, 2019; Lavy et al., 2012; Burke et al., 2013). Given the fact that the SES profile of a school is positively correlated with students' outcomes, it will also be positively related with the average performance level of schools as in more advantaged schools, students perform better than in less advantaged schools.

Less evidence exist regarding the relationship between a school's SES profile and the level of socioeconomic and language-based inequity. Only a few studies have analyzed whether or not low SES and non-native students are more sensitive to the effect of school SES then high SES and native students. Although it is often assumed that disadvantaged students are more sensitive to a school's socioeconomic profile, the little evidence that does exist, suggests otherwise. Willms (1986) observed for example that, in Scotland, the effect of school SES has just as much impact for high-SES students as it does for low-SES students. Similar results are found by Rumberger and Palardy (2006) for the United states, by Perry and McConney (2010a, 2010b, 2013) who analyzed this relationship in Australia and Canada and by Ammermueller and Pischke (2009) who examined this effect in Germany, France, Iceland, the Netherlands, Norway and Sweden. Furthermore, some studies also found the opposite effect than the one that is most often assumed: Palardy (2008) observed that in the US, more advantaged students actually derive more benefit from attending higher SES schools then disadvantaged students. Wells (2010) found similar results between immigrant and non-immigrant students. Finally, Cumberworth et al. (2018) also observed a positive interaction between school SES and student SES meaning that high-SES students benefit *more* from going to high SES schools than low-SES students. However, they assume that this positive interaction effect is largely explained by measurement error and needs to be interpreted very carefully.

[4] Note that PISA does not control for pre-existing differences between students.

2.2.2. The ethnic profile of a school

The relationship between the ethnic profile of a school and student performance is of growing interest in many countries. Although there is less consensus when it comes to the impact of ethnic school composition on student performance and equity, the matter is becoming more urgent given the increasing multiculturalism and multilingualism in education (OECD, 2005; OECD, 2012; Agirdag et al., 2012).

The main concern about the concentration of immigrant and non-native students in certain schools is its potential association with poorer student outcomes and consequently the increasing educational inequity, and of its adverse impact on upward social mobility (OECD, 2016A; Jensen, 2011). With respect to the relation between the ethnic composition of schools and student performance, research is inconclusive. While some authors suggest that a higher ethnic share – which refers to the proportion of migrant children in a school – is related to lower academic performance (Driessen, 2002; Dumay et al., 2008), others do not find any significant relationship between ethnic school composition and academic achievement, in particular when individual socio-economic status (SES), ethnic background and previous academic achievement are taken into account (Van der Slik et al., 2006; Fekjaer et al., 2007). To date, most studies used migration status to measure the effect of the ethnic profile of a school. Only a few studies focused on the share of non-native speakers on student performance and observed weak negative or no correlations at all with academic achievement (Tonello, 2016; Gaey et al., 2013).

Almost no direct evidence exists regarding the relationship between a school's ethnic profile and the level of socioeconomic and language-based inequity. However, it has been repeatedly demonstrated that immigrants and non-native students often go to schools with a higher percentage of immigrant and/or nonnative students and that are more disadvantaged in terms of SES (OECD, 2016b). Consequently, one can expect similar results to those concerning the relationship between student-inequity and the school SES profile. In this study, we hope to get a better view on how the proportion of non-native speakers affects student-inequity.

2.2.3. Percentage of certified teachers as a measure of the quality of teachers

Teacher quality is a much debated topic. It is well-known that the quality of teachers is an important factor influencing students' performance and student-inequity. Not only do they guide students towards their goals and shape their perceptions, they also have long-term effects on students' academic achievement, motivation, well-being and workforce capabilities (Caena, 2011; European Commission/EACEA 2015; Nusche et al., 2015; OECD 2005; OECD, 2017b; McKinsey et.al, 2007).

However, the teaching profession is losing its attractiveness. About half of the European countries report serious concerns about a shrinking supply of qualified teachers and almost all countries have concerns about attracting teachers that are "high achievers" (European Commission/EACEA/Eurydice, 2015; OECD, 2005). The fact that many countries face shortages of qualified teachers could have serious consequences on student outcomes. An extensive amount of research suggests for instance that teacher quality is positively linked with student learning and student performance (Clotfelter, et al., 2007; Clotfelter et al., 2010; Darling-Hammond, 2004; Monk, 1994; Ronfeldt et al., 2012; Caena, 2011; OECD, 2005) and that students who are not exposed to good quality teachers stand very little chance of progress (McKinsey; 2007).

Given the important role of teacher quality, some studies also focused on the relationship between teacher quality and student-inequity. The OECD observed for example that the quality of teachers can diminish the achievement gap between student from disadvantaged and advantaged schools (OECD, 2018b; OECD, 2013A; OECD, 2013b). Results are however not straightforward. The study of Luschei and Chudgar (2011), in which TIMSS 2003 data from 25 economically diverse countries are used, suggest for example that teacher characteristics (amongst which teacher quality) are equally important for students of different SES in primary education. By contrast, Chiu (2015) found that socioeconomic inequity is larger in schools with a higher share of qualified teachers (in terms of university degrees), by using pooled PISA 2009 data from 65 countries. A potential explanation for these findings could be the unequal distribution of qualified teachers amongst advantaged and disadvantaged schools. A plethora of research has demonstrated that in many countries, disadvantaged schools tend to have more difficulties in attracting qualified (in terms of educational certificates) and experienced teachers. For instance, in the Netherlands, the proportion of qualified teachers is three times higher in advantaged schools than disadvantaged schools (OECD, 2013a), and in the Flemish community, Sweden and Alberta, experienced teachers usually work in advantaged schools whereas teachers with less experience mostly work in disadvantaged schools (OECD, 2013a, 2017c). This unequal distribution of qualified teachers not only eliminates the positive link between teacher quality and equity, but even constitutes a major threat to the level of equity in education systems. In a recent PISA report (2017b), it has been revealed that the wider the gap between advantaged schools and disadvantaged schools in their science teachers' qualifications, the wider the difference in science performance between students in the top and bottom quarters of socio-economic status. However, this explanation is not in line with the findings of the study of Akiba et al. (2007). They used TIMSS 2003 grade 8 data from 46 countries to compare educational systems with needs-based[5] and unequal access to qualified teachers. Teacher quality was measured by several indicators (teaching certificates, mathematics as major, mathematics education as major, and teaching experience). The analyses reveal that the SES achievement gaps are mostly unrelated to the distribution of teaching quality. The only exception was observed for access to teachers with a major in mathematics; the SES achievement gaps were higher in countries where advantaged children have greater access to teachers with a major in mathematics.

2.2.4. Student-teacher ratio

The student-teacher ratio is closely intertwined with the previous school-level characteristic (teacher quality). The literature regarding the impact of the student-teacher ratio on student performance and the level of educational equity is contradictory. Some studies show that a smaller class size – and thus a lower student-teacher ratio – leads to better student performance and that this seems to benefit disadvantaged students (in terms of SES, migration background and mother tongue) more than others. Hence, the level of student-inequity in education would be lower by lower student-teacher ratios (Krueger et al., 2001; Ding et al., 2010; Shin, 2012; Bouguen et al., 2017; Bjorklund et al., 2005; Andersson, 2007; Nusche, 2009). Other studies do not find any impact of smaller student-teacher ratios, or even observe an adverse effect on students' performance and student-inequity (Leuven et al., 2008; Li et al., 2016; Shen et al., 2017; Ecalle et al., 2006; Jahanshahi et al.,. An OECD report (2018)

[5] In needs-based systems, disadvantaged children have greater access to qualified teachers.

showed for example that in countries that compensate for student disadvantage by reducing class size, the achievement gaps are, on average, not smaller or larger, compared to countries where class size is not related to students' SES, or where classes are larger in disadvantaged schools than in advantaged schools.

2.3. System-level characteristics

Apart from student- and school-level characteristics, system-level characteristics are also important determinants of educational success and inequity. Based on an extended literature review, we selected five system-level characteristics that could help us explain why some education systems succeed better in increasing performance and educational equity than others.

2.3.1. GDP per capita

GDP per capita is a commonly used context variable to measure the economic wealth of a country. Although many studies that analysed educational equity included a variable that reflects the economic wealth of an education system, it was often just used as a control variable. Hence, little attention has been devoted to the correlation between a country's level of economic wealth and its level of educational success and equity.

Despite the limited literature, studies that did include the measure, often observed a (weak) positive correlation between the economic wealth of a country and the academic performance of students and schools (Perry, 2009; Pfeffer, 2009; Hout and Diprete, 2006). This positive correlation could be interpreted in several ways; On the one hand, more economic wealth in a country raises the demand for a higher educated workforce. Consequently, parents will be more optimistic about the expected returns to education for their children, and therefore support and motivate their children better during their school career. This could lead to higher student performance and hence higher average school performance. Moreover, irrespective of the country's expenditure on formal education, a more wealthy environment facilitates informal learning (outside school) through cultural and recreational activities, and reduces the opportunity cost of learning. On the other hand, in the long run, higher educational performance in turn leads to increased human capital, which eventually results in more economic growth (Gylfason, 2003; Woessmann, 2008; OECD, 2012c).

Direct evidence examining the correlation between a country's level of economic wealth and the level of inequity, is missing. As mentioned above, the economic wealth of an education system is often used as a control variable.

2.3.2. School choice

A second system-level factor that is correlated with students' and schools' cognitive outcomes, and the level of educational inequity, is the degree of 'school choice'. This refers to the degree to which (parents of) students can decide which schools they will attend based on their preferences. In some countries, students are for example obliged to choose a school within a certain area, or the government assigns students to schools. Other countries leave the choice to the students (and parents) themselves (European commission/EACAE/Eurydice, 2014). Over the past decades, school choice programs have expanded in scope and size in many countries. More than two-thirds of the OECD countries have increased school choice opportunities for students since the 1980s (Musset, 2012; OECD, 2019B). Whether or not school choice has a positive or negative association with performance levels and the level of educational equity is, however, a widely debated topic (Montes, 2012; OECD, 2017E).

Proponents of school choice emphasize that parents have a general right to educate their children according to their personal preferences (in terms of religion or lifestyle) and hence have the right to choose the school that best matches these child-rearing preferences. Moreover, they argue that higher levels of school choice gives schools the right incentives to improve and innovate their education as this creates a more competitive market that forces schools to compete for 'students'. Consequently, school choice should improve student performance (OECD, 2019; Montes, 2012; OECD, 2017E; Betts and Loveless, 2005; Hoxby, 2000). Empirical evidence on whether academic achievement is actually higher when there is greater competition between schools is mixed, however (Boeskens, 2016; Urquiola, 2016). While some studies conclude that school choice and school competition enhance the performance of students and schools (Epple et al., 2017; Hoxby, 2002), other analyses show more disappointing results (OECD, 2019; Abdulkadiroglu, et al.,2018).

By contrast, opponents argue that school choice increases the sorting of students (or segregation) between schools, based on students' characteristics such as their SES or their ethnic background, which eventually only exacerbates inequalities between students (and schools). From this point of view, free school choice involves a trade-off between efficiency and equity objectives. A large body of research supports this hypothesis and showed that school choice does indeed increase student- and school-inequities (Ladd and Walsh, 2002; Ladd, 2003; Soderstrom and Uusitalo, 2005; Field et al, 2007; Ladd, et al., 2011; Musset, 2012; Cullen, 2005; Garcia, 2008; Böhlmark, 2007).

Altogether, studies revealed a clear negative relationship between the degree of school choice in an education system and student- and school-inequity, whereas its association with students' and schools' academic achievement is more inconclusive.

2.3.3. The practice of grade repetition

Ample evidence proved that grade repetition is a costly and ineffective practice for both individuals and the society (OECD, 2012d). It has a negative impact on student educational outcomes and increases inequity as disadvantaged students are more likely to fall a year behind.

Grade repetition occurs when students, after a formal or informal assessment, are held back in the same grade for an additional year. Teachers widely support the practice as they observe some benefits of repetition, which are related to immediate gains when going through the same curriculum a second time (Jimerson, et al., 2002). However, research showed that benefits are only slight and short-lived (OECD, 2012d; Brophy, 2006). In the long-term, it has negative financial, social and academic effects. Despite these negative sides of grade repetition, it is still practiced on a fairly large scale in many OECD countries: in 2015, 12% of all 15-year-olds reported to have repeated at least one year either in primary or secondary school, on average amongst all OECD countries (OECD, 2016B). In countries where grade repetition is extensively used – such as France, Luxembourg, Spain, Portugal and Belgium – it affects over 30% of the students.

First, with respect to the negative financial effects of grade repetition, a study of Jimerson, Pletcher and Graydon (2006) revealed that the direct costs of grade repetition are very high for education systems, as these include providing an additional year of education and delaying entry to the labour market by a year. Moreover, as students who repeat a year are more likely to engage in high-risk behaviour and/or dropout, repetition increases expenditure on other social services (Jimerson et al., 2006). Second, regarding the academic impact of grade repetition, studies observed that school systems that extensively use grade repetition are associated with lower student performance. Moreover, it increases the likelihood of earning no qualification or only a lower secondary one (Jacob and Lefgren, 2009). Third, grade repetition also has a negative psycho-social impact on students. More specifically, grade repetition is a source of stress, ridicule and bullying by others (Yamamoto and Byrnes, 1987; Anderson, Jimerson and Whipple, 2005), which negatively affects self-esteem and increases the likelihood of high-risk behaviour, school failure and dropout, as noted above.

Not all students have, however, the same chance of repeating a grade. Students with low socio-economic backgrounds, poorly educated parents or immigrant backgrounds, and boys, are significantly more likely to repeat a grade than others (OECD, 2011b). Consequently, grade repetition is associated with higher levels of student- and school-inequity as the achievement gap between disadvantaged (in terms of SES and ethnicity) and advantaged students, enlarges. Education systems that more frequently use the practice of grade repetition will show higher levels of socioeconomic and language-based inequity.

2.3.4. Tracking age

A fourth well-known system-level characteristic that is correlated with student performance and especially the level of educational equity is the age of first 'tracking' of students (often referred to as the practice of 'early tracking'). Tracking refers to the practice whereby pupils with different perceived abilities are oriented through distinct curricula towards different educational end-points. A frequent argument against tracking is that students are (often) not solely sorted into tracks based on their abilities. Their SES also plays a (major) role. Studies proved for example that disadvantaged students (both in terms of SES and linguistic background) are more often placed in less intensive tracks, than their more advantaged counterparts, even after controlling for performance (Woesmann, 2009). Consequently, disadvantaged students will be frequently overrepresented in lower tracks, ending up more often in ghetto schools and getting demotivated. Given the fact that 'lower' tracks are less challenging than 'higher' tracks, tracking is seen as harmful for students' performance, in particular for disadvantaged students. From this point of view, many studies argued that the age at which students are tracked can be of great importance for student performance and educational equity. More precisely, when students are tracked at younger ages, their 'choices' will be heavily influenced by their role models, such as their parents. In that sense, it can be expected that students' socioeconomic background is more of a burden when tracking decisions are made at younger ages (Lavrijsen & Nicaise, 2015; 2016). Many studies have validated this observation. The study of Dupriez and Dumay (2006), in which primary and secondary school data for 15 European countries that participated in both PIRLS 2001 and PISA 2000 are used, revealed a mitigating effect of the age of tracking on inequity. By combining data of primary education with data of secondary education in a 'pseudo-panel' framework, one can control for initial levels of inequality. Similar results are observed in the studies of Lavrijsen & Nicaise (2015; 2016), in which data from more than 30 countries were analyzed. Finally, higher levels of inequity in early tracking countries were also found in studies that did not control for initial levels of inequity (Schütz et al., 2008; Brunello & Checchi, 2007; Horn, 2009; Le Donné, 2014; Schlicht et al., 2010).

Far less evidence is available on the presumed equity-efficiency dilemma in the tracking debate. One of the only (?) studies that explicitly deals with this issue, is the study by Lavrijsen & Nicaise (2015; 2016). Their pseudo-panel study used a dif-in-dif design to compare the relative gains or losses of various subgroups of students between primary and secondary, in early versus late tracking systems. In each dataset, students were ranked into percentiles based on their performance in a given subject. Thus, it was possible not only to estimate the average gain or loss from later tracking, but also the gain or loss for any quantile in the distribution separately. The study found

(1) that, across the board, late tracking never results in average losses in performance. This applies to all three subjects examined (maths, science, reading), irrespective of the pair of primary and secondary level datasets selected. In reading there are even significant average gains from late tracking;
(2) that the largest gains from late tracking are made by the weakest / most disadvantaged students, especially in reading, whereas the gains for the strongest groups are small and insignificant;
(3) for maths only, the 25% top performers appear to lose from late tracking, but this loss is not statistically significant.

We can conclude that the evidence on the relationship between the tracking age and the level of equity is conclusive and shows a positive correlation. In other words, the earlier the school tracking, the greater the impact of family background on educational success and thus the lower the level of educational equity. The later students are tracked into different tracks, the more equitable the education system. Moreover, late tracking seems to result in upward rather than downward levelling of cognitive outcomes. In other words, equity and efficiency arguments go hand in hand, in favour of postponed tracking.

2.3.5. Government expenditure on primary education (as a % of GDP)

The last system-level characteristic included in this report is the expenditure on primary education by the government. The seemingly straightforward question - whether or not an increase in government education funding yields better educational outcomes and improves equity - has produced considerable controversy during the past decades.

With respect to the first part of the question – whether government expenditure on education is positively correlated with average educational outcomes – the empirical findings are inconclusive. Some studies state that the amount spent per pupil does matter and is positively associated with student (and school) performance. In a recent study of the Boston Consultation Group, where students of the 4th and 8th grade were examined between 2003 and 2015 in 50 states of the US, it was demonstrated that government spending per student was positively correlated with students' performance, in particular for disadvantaged students (Gjaja, 2014). This conclusion is supported by Jackson et al. (2007): the authors evaluated the K-12 program in the US and found that a 10% increase in per pupil spending each year for all 12 years of public school, is associated with 0.46 additional years of completed education, 9.6% higher earnings, and a 6.1 percentage point reduction in the annual incidence of adult poverty. Moreover, several studies revealed that if the increased funding is spent efficiently, it can (and often does) influence student achievement positively (An, 2013; Matthew et al. 2014; Struhl et al., 2012; Berger et al., 2013; Edmunds et al., 2017; US Department). For example, spending money on dual enrollment programs has many significant positive effects on high school outcomes —such as increased attendance, academic achievement, and completion—as well as college outcomes — such as increased readiness, access and enrollment, credit accumulation, and attainment. In the same vein, spending on early childhood programs such as Head Start 21, the Perry Preschool Project or the Carolina Abecedarian Project also proved to be positively linked with student achievement at later ages. By contrast, other researchers do not find any relationship between government expenditure on education and student outcomes (OECD, 2012c; Okpala, 2002; Tajalli, 2005). Hanushek's research (2003, 2006) argues for instance that higher spending on inputs (such as lower student-teacher ratios, the percentage of teachers with a master's degree, and per-pupil expenditure) has risen in the United States for the past five decades, without commensurate improvements in student achievement. Hence, higher government spending – regardless of the kind of investment – on education does not appear to yield better student performance. Furthermore, The OECD (2012c) states that when the cumulative public expenditure per student from age 6 to age 15 exceeds 35,000 USD, the significant positive relationship with student performance disappears. In sum, to say the least, it is not clear whether higher government expenditure on education is correlated with better student achievement. It can be assumed that the way money is spent (efficiency) is a more important factor with regards to increasing student achievement than the volume of spending.

A clearer picture emerges regarding the correlation between government expenditure on education and inequity. International studies (such as TIMSS and PISA) repeatedly observe that higher government expenditure is associated with higher levels of student- and school-inequity (Strietholt, 2019). This positive association is observed in studies analyzing European countries (Schlicht et al., 2010), Middle East and North African countries (Salehi-sfahani et al., 2013), but also in analyses where a diverse sample of countries around the world is used (Akiba et al., 2007; Bodovski et al., 2017).

In this report, we specifically included government expenditure on *primary* education only. The reasoning behind this is twofold. First, we did some preliminary analyses in which the average level of educational equity in an education system was plotted against government expenditure on primary, secondary and tertiary education. Only for primary education, we observed a strong positive association. Second, research has pointed out that the efficiency of educational interventions for disadvantaged groups is inversely related with their age: very high for infants and toddlers, high at primary level, and rather modest at secondary level and beyond (Heckman, 2008; Cuhna et al., 2006). Our assumption is that the causal pattern is more complex at secondary level: whereas higher expenditure should normally boost the quality of education, it may also be due to inefficient fragmentation in tracked systems.

3. Research questions, data and method

3.1. Research questions

We propose four research questions:

RQ1 Which school- and system-level characteristics are correlated with the *mathematics and reading performance* of 15-year-old students and how?

RQ2 Which school- and system-level characteristics are correlated with *socioeconomic and language-based student-inequity* and to what degree does socioeconomic and language-based student-equity vary with these characteristics?

RQ3 Which school- and system-level characteristics are correlated with the average mathematics and reading performance of all 15-year-old students *at school level* and how?

RQ4 Which system-level characteristics are correlated with *socioeconomic and language-based school-inequity* and how?

3.2. Data

Data of the 6th wave (2015) of the Program for International Student Assessment (PISA) are primarily used in this report. PISA is an international triennial test of student achievement organized by the Origanization for Economic and Development (OECD) since 2000. According to the OECD, "PISA assesses the extent to which 15-year-old students, near the end of their compulsory education, have acquired key knowledge and skills that are essential for full participation in modern societies." (OECD, 2017F, p.12). Students are assessed in three main domains in each round: science, mathematics and reading. One of these domains is selected in each wave as the major domain and is given greater emphasis than the remaining two domains. In 2015, science was the major domain, yet the other domains can just as well be analyzed and compared over time.

In 2015, 72 countries participated to the survey – 34 OECD countries and 38 partner countries – from which 50 countries are included in this study. Not all countries were included due to: (1) a lack of sufficient data regarding certain variables in some countries and, (2) our focus on developed countries. After the selection procedure, our sample consisted of 301,998 students in 11,368 schools.

However, the PISA 2015 data did not include data on most of our system-level characteristics. Therefore, we borrowed data from several other sources (Eurydice, World Bank, PISA 2012) and added them to the PISA 2015 data. Each of the indicators and their sources will be discussed in detail in section 3.2.1.2.

In what follows, we will briefly elaborate on the complexity of using PISA data. More detailed information can be found in the many technical reports of the OECD.

Plausible values

In many large-scale assessment programs (amongst which PISA), students' cognitive outcomes are measured by plausible values. Plausible values are multiple imputations of the unobservable latent achievement for each student (Wu, 2005). Although it is often recommended to use all available plausible values, we only used one plausible value in this study. The reasoning behind this decision is threefold: (1) models that were estimated with just 1 plausible value already necessitated a lot of computing time, (2) the OECD stated that "using one plausible value or five does not really make a substantial difference on large samples" (OECD, 2009, p.44).

Study-design

PISA uses a complex two-stage sampling design. In a nutshell, within each participating country, a random sample of schools is selected, and then a random sample of about 42 (=target cluster size) 15-year-old students is selected with equal probability within each school. In other words, PISA has a nested structure in which students are nested into schools into countries. This nested structure needs to be taken into account to avoid violating the assumptions of homogeneity of clusters, uncorrelated error terms and independence of observations. Moreover, by ignoring the nested structure we would underestimate the population variance, leading to a type I-error.

If schools had fewer 15-year-old students than the target cluster size of 42 students, all students on the list were selected (Cumberworth, 2018; OECD, 2016c). This implies, however, that some schools only have a few sampled students (> 10 students). Using individual-level variables and aggregating them at school-level in order to measure school-level characteristics could in this case generate less reliable estimates. However, all models were estimated with and without these small-group schools, and we found that excluding them did not affect our results. Rather than choosing an arbitrary cut-off point, we present results with all schools included.

Finally, another important aspect concerning the study design is the use of survey weights. Weights are required to analyze PISA data, to calculate appropriate estimates of sampling error and to make valid estimates and inferences about the population. Weights must be included into the analysis to ensure that each sampled student and school appropriately represents the correct number of students and schools in the full PISA population. In this study we used all necessary weights for that purpose.

Shortcomings of PISA

The study design is –as with many datasets – not without flaws and there are some shortcomings in particular to which we must give careful attention: (1) PISA is a cross-sectional survey and (2) long-term outcomes are not measured in PISA.

First, PISA is a cross-sectional survey with no measure of student ability prior to entering the current school. Consequently, results are not corrected for unequal 'starting positions', i.e. when students access their school in the first grade. Without this correction, one could argue that students are not randomly assigned to schools, as the beta-coefficients that are observed at the age of 15 capture the influence of antecedents in the school career that are correlated with their social background. However, our research focuses on school- and system-effects throughout the school career: it aims to capture the *cumulative* impact of social inequalities on cognitive skills at age 15, whatever their origin (social capital acquired in the students' home environment as well as sorting during the entire past school career) and whatever the stage at which they were shaped. Furthermore, using a cross-sectional survey means that causality cannot be proven. Hence, all results should be interpreted as correlations, rather than causal effects.

Second, PISA contains no measure on long-term outcomes. Test scores may be good predictors of later achievement, but certainly are not perfect. For example, there is some evidence that school composition effects have a bigger impact on long-term outcomes than on standardized test scores (Rumberger and Palardy, 2005b).

3.2.1. Variables

3.2.1.1. Dependent variables

Student-level

Mathematical literacy: the OECD defines Mathematical literacy as “students’ ability to analyze, reason and communicate ideas effectively as they pose, formulate, solve and interpret solutions to mathematical problems in a variety of situations” (OECD, 2016a, P27).

Reading literacy: the OECD defines reading literacy as “students’ ability to understand, use, reflect on and engage with written texts in order to achieve one’s goals, develop one’s knowledge and potential, and participate in society” (OECD, 2016a, P27).

School-level

School mathematical literacy: this variable reflects the mathematical literacy of a school. It was constructed by taking the average of students’ mathematical literacy in each school.

School reading literacy: this variable reflects the reading literacy of a school. It was constructed by taking the average of students’ reading literacy in each school.

3.2.1.2. Independent variables

Control variables

Age: the age of a student is a scale computed by the OECD. It was calculated as the difference between the year and month of the testing and the year and month of a student’s birth. In other words, the higher the scale, the older a student.

Gender: gender is a nominal variable created by the OECD with two categories: (1) boys, (2) girls. We dummified this variable with girls as the reference category.

Migration status: to measure the migration status of a student, the OECD created a nominal variable with 3 categories: (1) native students (those students who had at least one parent born in the country), (2) second-generation students (those born in the country of assessment but whose parent(s) were born in another country) and (3) first-generation students (those students born outside the country of assessment and whose parents were also born in another country). We dummified this variable, with native students as the reference category.

Student-level characteristics

Socioeconomic status (SES): to measure the individual SES of students, the index of economic, social and cultural status (ESCS) – a scale constructed by the OECD – is used. The ESCS is a composite score built via principal component analysis combining three indicators; (1) parental education, (2) highest parental occupation, (3) and home possessions including books in the home. The scale is standardized over all OECD countries with a mean of 0 and a standard error of 1. This deviates slightly when adding non-OECD countries to the sample.

Language spoken at home: PISA has an international comparable variable that indicates whether or not the assessment test is in the same language than the one that students speak at home. This nominal variable has two categories: (1) students' language at home is the same as the language of assessment for that student (native speakers) and (2) students' language at home is another language (non-native speakers). In this study, we dummified this variable with native speakers being the reference category.

School-level characteristics

Socioeconomic school profile: this variable was constructed by taking the average of students' SES in each school.

Concentration of students speaking another language at school: similarly, this variable was constructed by taking the percentage of students that speak another language at home than the one spoken at school.

Percentage of certified teachers: the percentage of fully certified teachers is a variable computed by the OECD by dividing the number of fully certified[6] teachers in a school by the total number of teachers in that same school.

Student-teacher ratio: the student-teacher ratio is also a variable computed by the OECD by dividing the number of enrolled students in a school by the total number of teachers (FTE) in that same school.

System-level characteristics

Gross Domestic Product per capita (GDP): the economic wealth of a country is measured by the GDP per capita of a country (current U.S. dollars) in 2015. As the variance of the variable was too large for MPlus to handle, we divided it by 10,000. Data are borrowed from World Bank (World Bank, 2015)

Age of first tracking: we borrowed data from the OECD and the Eurydice report 'the structure of European education systems – 2014/2015' to create a variable that reflects the age at which students are tracked for the first time in an education system (Eurydice, 2015; OECD, 2015). The age range of the variable was 10 to 16 years. Consequently, we decided to dummify this variable with '0' being students that are tracked before the age of 15, and '1' being students being tracked from the age of 15 or later.

Government expenditure on primary education: to measure the expenditure on primary education, we used the indicator 'Government expenditure per student in primary education, as a percentage of GDP per capita in 2009'. As students in PISA are 15 years old, we included a time lag of 6 years. Data are borrowed from the World Bank (World Bank, 2009).

School choice: school choice is a complex concept to measure. As there are many forms in which school choice appears, proxies are commonly used. In several PISA-reports where school choice is discussed, the OECD used the competitiveness between schools as a proxy for school choice (OECD, 2019).

[6] Fully certified by the appropriate authority

Competition between schools is after all (often) a consequence of school choice. In this study, similarly, the competition between schools will be used as measure for the level of school choice in an education system.

The competitiveness between schools is measured in the PISA-survey by asking principals about the number of schools in their neighbourhood that are competing for students. The variable had three categories: (1) no schools, (2) one school, (3) two or more schools. A principal could only check one box. Based on this variable, we constructed a competitiveness index that reflected the percentage of competition between schools on average in an education system. The index was created with the following formula: $[(\alpha + 2*\delta)/3]$. In this formula, 'α' refers to the percentage of principals in an education system that checked category two, indicating that there is one other school in the neighbourhood competing for the same students. 'δ' refers to the percentage of principals in an education system that checked category three, indicating that there are two or more schools in their neighbourhood competing for the same students. We gave δ a double weight. Finally, we divided the sum by three to obtain a value between 0 and 100.

Repetition rate: the repetition rate reflects the percentage of students that have already repeated one or more years in primary or secondary education by the age of 15. In the PISA survey, students were asked whether or not they had repeated one or more grade). The system-level variable is the percentage of students in an education system who checked that box.

Table 1: Descriptive statistics

	Interval variables					
	Mean	**SD**	**Min**	**Max**	**5th percentile**	**95th percentile**
Age	15.79	0.29	15.17	16.42	15.33	16.27
Student SES	-0.05	0.98	-7.05	4.18	-1.69	1.38
% Certified teachers	0.87	0.27	0.00	1.00	0.06	1.00
Student-Teacher ratio	12.84	5.42	1.00	100	6.00	20.33
School SES	-0.05	0.62	-3.82	1.58	-1.12	0.88
School % non-native speakers	12.86	20.89	0.00	100.00	0.00	64.71
Expenditure primary education	21.01	5.07	10.59	41.72	12.00	27.95
Competitiveness between schools	46.69	11.59	17.56	69.09	22.22	64.19
% students that have repeated a grade	11.12	12.26	1.07	69.71	1.49	36.51
GDP per capita/10000	3.41	2.18	0.22	10.04	0.42	7.53
	Categorical variables					
	%					
Language						
Native speakers (=0)	87.20					
Non-native speakers (=1)	12.80					
Migration status						
Non-immigrant student (=0)	87.60					
Immigrant student (=1)	12.40					
Gender						
Female (=0)	50.20					
Male (=1)	49.80					
Tracking age						
Before 15 years old (=0)	37.00					
15 years old or later (=1)	63.00					

3.3. Method

Given the nested structure of our data, hierarchical linear modelling is our primary method. SAS 9.4 was used to prepare data and Mplus version 8 was relied upon for the statistical analyses. Missing data were handled by the full information maximum likelihood procedure, which is implemented in Mplus. In all analyses, we used the robust maximum likelihood estimation. Finally, all variables – excepting dummy variables and our dependent variables – in our models are grand mean centered.

3.3.1. Models

As we have dependent variables at student- and school-level, we included both three-level and two-level models in this study. Three-level models were used when we modelled the dependent variables at student-level. Two-level models were used when we modelled the dependent variables at school-level.

Our initial idea was to estimate three-level models, in which students (L1) are nested into schools (L2) into education system (L3). At each level, predictors would be included and cross-level interaction terms between levels would be added. However, due to the complexity of these models, many statistical problems came up during the estimation of the models (for example non-convergence issues, too many parameters, etc.). Consequently, we decided to split the three-level model into several models where we only included predictors on two levels to reduce the complexity, while keeping the hierarchical structure of a three-level model. In other words, when analysing for example how system-level factors are correlated with student performance and student-inequity, we nested students into schools into countries, but only included student- and system-level predictors. At the level where no predictors were included (school-level in this case), the random intercept and slopes were fixed. In the two-level models, schools (L2) are nested into education systems (L3) and predictors at both levels are always included.

3.3.2. Model construction

All models were built in the same way. In Models A we included a random intercept and fixed parameters. Hence, we could already draw some conclusions about the correlation between school- and system-level characteristics with students' and schools' academic performance. In Models B, random slopes of students' SES and language spoken at home were added. By including these random slopes, we could test whether or not the relationship between students' and schools' SES and/or language-based background varied between schools and/or education systems. Finally, in Models C, we added cross-level interactions between the random slopes and the included predictors at school- or system-level. These cross-level interactions showed whether or not student- and school-inequity varied with school- and/or system-level characteristics.

We performed six analyses (three for mathematics performance, three for reading performance). In the first analyses (see Figure 1), three-level models were estimated with student-level variables (L1) and school-level variables (L2). The third level did not contain any predictors and the random intercept and slopes were fixed. In the second analyses (see Figure 2), we estimated again three-level models, but this time student-level variables (L1) and system-level variables (L3) were included. Consequently, the second level had no predictors and the random intercept and slopes were fixed. In the last analyses (see Figure 3), we estimated two-level models in which schools (L2) are nested within education systems (L3). Predictors on both levels are included.

An overview of all the estimated models is presented in Table 2. Next, the C-variants of the multilevel models – for mathematics – are presented graphically. Note that cross-level interactions were estimated only with SES and language-based origin.

Table 2: Overview of all models estimated

Analysis 1: L1 nested into L2 – Mathematics	➢ Model 1A ➢ Model 1B ➢ Model 1C
Analysis 2: L1 nested into L2 - Reading	➢ Model 2A ➢ Model 2B ➢ Model 2C[7]
Analysis 3: L1 nested into L3 - Mathematics	➢ Model 3A ➢ Model 3B ➢ Model 3C
Analysis 4: L1 nested into L3 - Reading	➢ Model 4A ➢ Model 4B ➢ Model 4C
Analysis 5: L2 nested into L3 – Mathematics	➢ Model 5A ➢ Model 5B ➢ Model 5C
Analysis 6: L2 nested into L3 - Reading	➢ Model 6A ➢ Model 6B ➢ Model 6C

[7] Model 2C– after reducing the complexity – still faced problems in converging. Consequently, we decided to lower the convergence criterium.

Figure 1: Simplified version of Model 1C: three-level model with student- and school-level characteristics

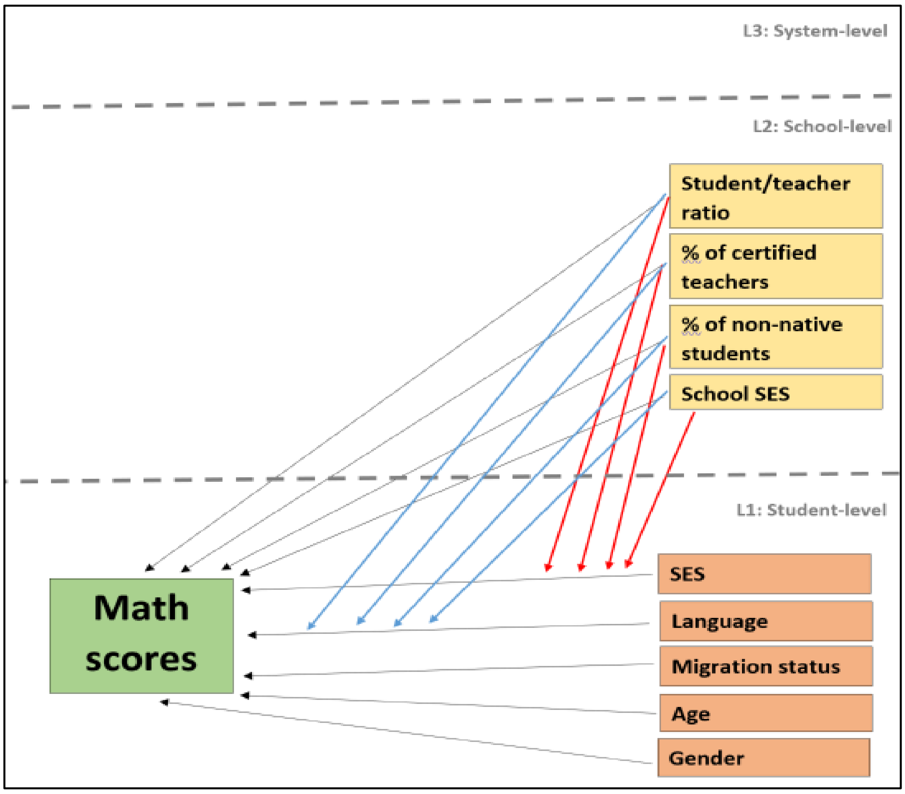


Figure 2: Simplified version of Model 3C: three-level model with student- and system-level characteristics

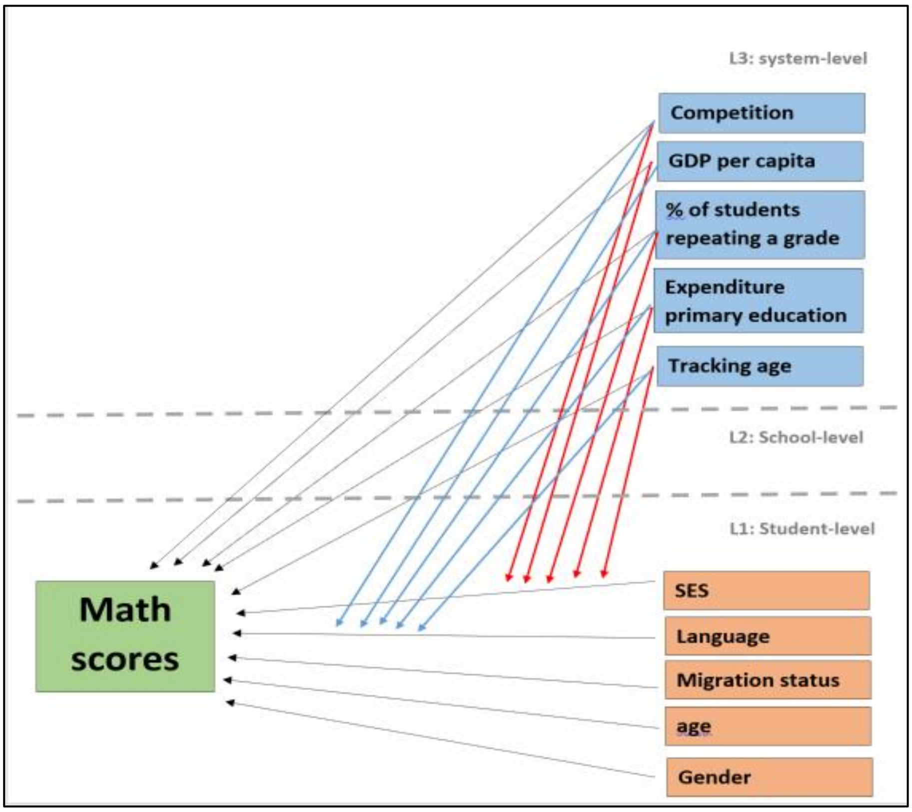
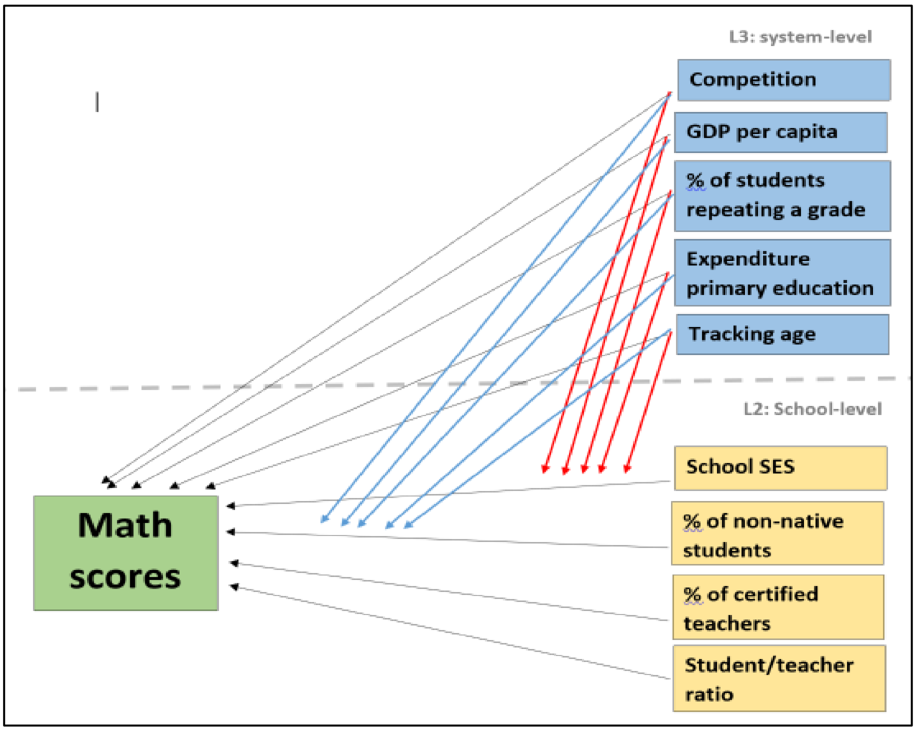


Figure 3: Simplified version of Model 5C: two-level model with school- and system-level characteristics



.

4. Results

The aim of our research is not only to examine which school- and system-level characteristics are correlated with student and/or school performance, but also to get a better understanding of which school- and system-level characteristics determine socioeconomic and language-based student- and school-inequity. As mentioned before, socioeconomic and language-based student-equity are measured by: (1) the SES-gradient which refers to the correlation between students' SES and their mathematics or reading performance, and (2) the language-gradient which refers to the correlation between students' language spoken at home (whether or not it is the same as the one spoken at school) and their mathematics or reading performance. Similarly, socioeconomic and language-based school-inequity is estimated by two measures: (2) the school SES-gradient, i.e. the correlation between a school's average SES and the average mathematics or reading performance of students in that school, and (2) the language-gradient, i.e. to the correlation between the percentage of students that speak another language at home than the instruction language at school and the average mathematics or reading performance of students in a school. To examine how these inequity-measures vary with school- and/or system-level characteristics, multiple cross-level interactions were added in the models. To avoid a plethora of tables and numbers only models C – in which all interaction terms are included – will be presented in this paper[8].

Before presenting the results, we will briefly discuss an important feature of multilevel modelling: the Intraclass Correlation Coefficient (ICC). The ICC allows us to see how much of the variance in student and/or school performance (the dependent variables) can be explained by differences between individuals, schools and education systems. For student mathematics performance, the ICC reveals that 24% of the variance in students' mathematics performance can be explained by differences between schools and 34% by differences between education systems. This shows the importance of taking into account both school- and system-level factors when analysing socioeconomic and language-based inequity. Differences between schools and education systems explain, altogether, 58% of the variance in student mathematics performance. A similar conclusion can be drawn regarding student reading performance: 25.5% of the variance in students' reading performance is explained by differences between schools, while 34.4% is explained by differences between education systems. With respect to schools' mathematics performance, the ICC tells us that 41.2% of the variance in school mathematics performance can be explained by system-level variables. This is again quite high and proves that system-level characteristics are important when examining inequalities between schools. The ICC for schools' reading performance is more or less the same: 39.1%. In other words, 39.1% of the differences in the reading performance of schools can be explained by differences between education systems.

Finally, our statistical analyses do not 'prove' a causal explanation due to the fact that PISA is a cross-sectional survey. Consequently, when the terms "explain", "impact", "effect" or "result" are

[8] All model estimations can be obtained upon request from the authors.

used in the upcoming 'Results' and 'Discussion' sections (§4, §5) they should be understood in a statistical rather than a causal relationship sense.

4.1. School-level characteristics, student performance and student-equity

The intercepts in both models (Table 3, Model 1C, 2C) reflect the average score of a non-immigrant, native speaking, female student with an average age (15.79), an average SES (-0.05), that attends a school with an average student-teacher ratio (12.84), an average school SES (-0.05), an average percentage of non-native speaking students (12.86%) and an average percentage of fully certified teachers (87%).

Table 3: school-level characteristics and student performance

	Model 1C: Maths		**Model 2C: Reading**	
	B	**SE**	**B**	**SE**
Intercept	486.73***		493.75***	
Level 1: student-level				
Age *(centered)*	9.47***	1.34	9.58***	1.30
Migration status *(0 = non-immigrant students)*	-4.33	2.99	-12.72***	4.26
Language *(0 = native students)*	-21.99***	2.32	-8.75*	4.13
Gender *(0 = female)*	10.05***	1.09	-22.84***	1.56
SES *(centered)*	16.52***	0.92	3.96†	2.04
Level 2: school-level				
% certified teacher *(centered)*	26.60***	6.16	28.28†	17.13
Student-Teacher ratio *(centered)*	-0.86***	0.03	-0.16	0.82
School SES *(centered)*	50.33***	0.45	76.51***	0.85
% non-native speakers at school *(centered)*	-0.31*	0.14	-0.63*	0.32
Random slope school-level				
SES slope	7.95***	0.05	2.71*	1.15
Language slope	22.10***	0.80	37.01***	0.96
Cross-level interactions				
M_SES*SES	5.76***	0.00	2.03***	0.08
M_LANG*SES	0.00	0.03	0.03	0.07
% CERT TEA*SES	0.25***	0.05	-2.23	3.08
ST/TE-ratio *SES	0.02***	0.00	0.08	0.05
M_SES*LANG	2.59***	0.09	-7.38***	0.12
M_LANG*LANG	0.25***	0.03	0.29†	0.16
% CERT TEA*LANG	6.64***	1.64	1.73	1.23
Student-teacher ratio*LANG	0.65***	0.00	0.12	0.16

***p<0.001, **p<0.01, *p<0.05, † p<0.1

Note that including interaction terms in a model, changes the interpretation of the main effects. As cross-level interaction terms between all school-level and student-level characteristics (SES and language spoken at home) are included in Table 3, the main effects of these characteristics need to be interpreted differently. Student-level characteristics should be interpreted as: the relation between students' SES or their language spoken at home and student performance, *in a school with an average percentage of fully certified teachers, an average school SES, an average student-teacher ratio and an average percentage of non-native students at school.* The main effects of all school-level characteristics *always only apply for native speaking students with an average SES.* How these school-level effects vary when examining non-native speaking students and/or students with a higher or lower SES, will be discussed in section 4.1.4. in which the cross-level interaction terms are examined.

4.1.1. Control variables

We will start by briefly discussing the correlations between our control variables (age, gender and migration status) and student performance. With respect to age, we observe a positive significant correlation with both mathematics (B=9.47) and reading (B=9.58) performance, implying that older students outperform their younger peers. Second, in both models migration status is negatively associated with student performance, but this correlation is significant only in Model 2C in which student reading performance is modelled. Hence, the reading performance of immigrant students is 12.72 points lower than those of non-immigrant students, whereas mathematics performance does not differ between immigrant and non-immigrant 15-year-olds. The absence of a significant association in Model 1C, is probably due to fact that students' SES and students' language spoken at home are also included in the models and already capture the disadvantage of being an immigrant student[9]. Third, the gender of a student is positively correlated with student mathematics performance (B=10.05), but negatively with student reading performance (B=-22.84). Hence, boys score better on maths tests than girls, while girls outperform boys in reading.

4.1.2. Student-level characteristics and student performance

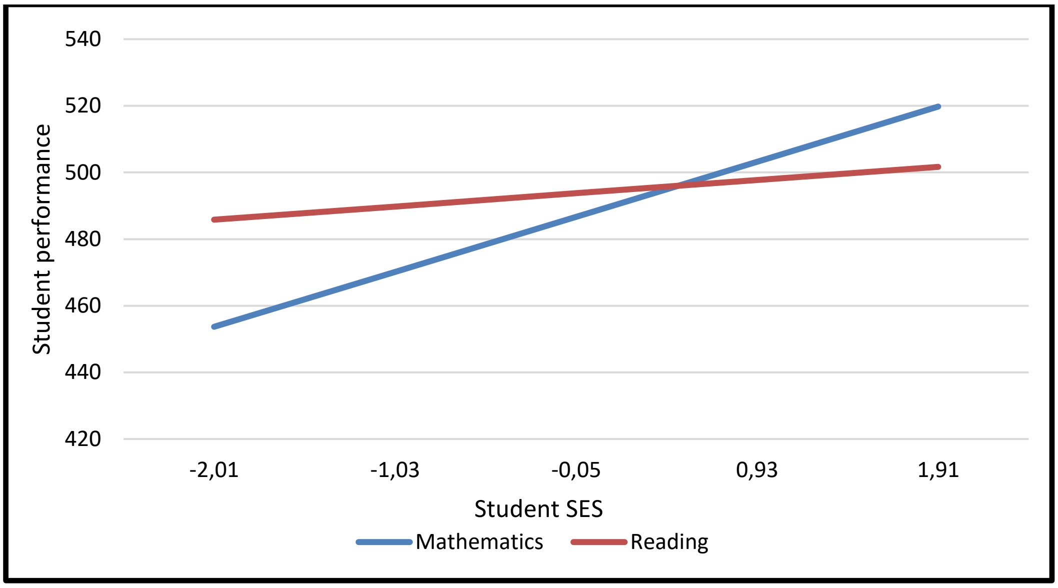
With respect to the correlation between student performance and language spoken at home, we find a negative significant correlation in both models. This means that – across all countries – 15-year-old non-native speakers perform worse in reading and mathematics than native speakers (Model 1C, B=-21.99; Model 2C, B=-8.75). The relation between language spoken at home and student mathematics performance is, however, stronger than the one with student reading performance. This difference could be explained by the strong negative interaction term between a students' language spoken at home and the SES of a school that can be observed in Model 2C, but not in Model 1C. We will elaborate more on this effect in section 4.1.4.

Concerning the correlation between students' SES and their performance, Table 3 reveals a positive significant correlation for both mathematics and reading performance (Model 1C, 2C). This indicates that – across all countries – 15-year-old students with a higher SES (1 standard deviation above the mean), score 16.52 points higher in mathematics and 3.96 points higher in reading than students with an average SES. Based on the descriptive table (Table 1) we can calculate the achievement gap in mathematics and reading between students with a very high SES (95th percentile) and students with a very low SES (5th percentile). Approximately, 1.67 standard deviations should be subtracted for students with a very low SES and 1.46 standard deviations should be added for students with a very high SES. Based on these standard deviations, the achievement gap between students with a high and low SES, is equal to 51.71 points in mathematics and 12.39 points in reading (see Figure 4). These

[9] Immigrant students are often disadvantaged and speak another language at home.

achievement gaps are equal to resp. more than 1 years of formal schooling and approx. 4 months of formal schooling[10].

Figure 4. Student mathematics and reading performance by student SES



4.1.3. School-level characteristics and student performance

Table 3 (Model 1C and 2C) reveals that almost all school-level characteristics have a significant correlation with mathematics and reading performance.

First, with respect to the association between the percentage of fully certified teachers[11] in a school and student performance, we observe in both models (Model 1C, 2C) a significant positive correlation of about the same effect size (Model 1C, B=26.60; Model 2C, B=28.28). This means that a 10 percentage points increase in the proportion of fully certified teachers in a school is associated with an increase of 9.85 points in students' mathematics performance, and an increase of 10.47 points in their reading performance. Second, the student-teacher ratio is negatively correlated with student performance in both models, but shows a significant correlation only with students' mathematics performance (Model 1C, B=-0.86). Hence, the higher the student-teacher ratio, the lower the mathematics performance of 15-year-olds. The correlation is, however, negligible: an additional 5.42 students in a class, leads to a decrease of 0.89 points in students' mathematics scores, which implies that the student-teacher ratio is not substantially related to student performance. Third, a large significant positive correlation between school SES and student mathematics (B= 50.33) and reading (B= 76.51) performance is revealed (Model 1C, 2C). This indicates that the school SES is a strong determinant of students' cognitive outcomes[12]. Similar to the calculation of the achievement gaps based on students' SES, we

[10] Keep in mind that these individual-level correlations are partly underestimated due to the measurement error that is captured by the school SES effect.

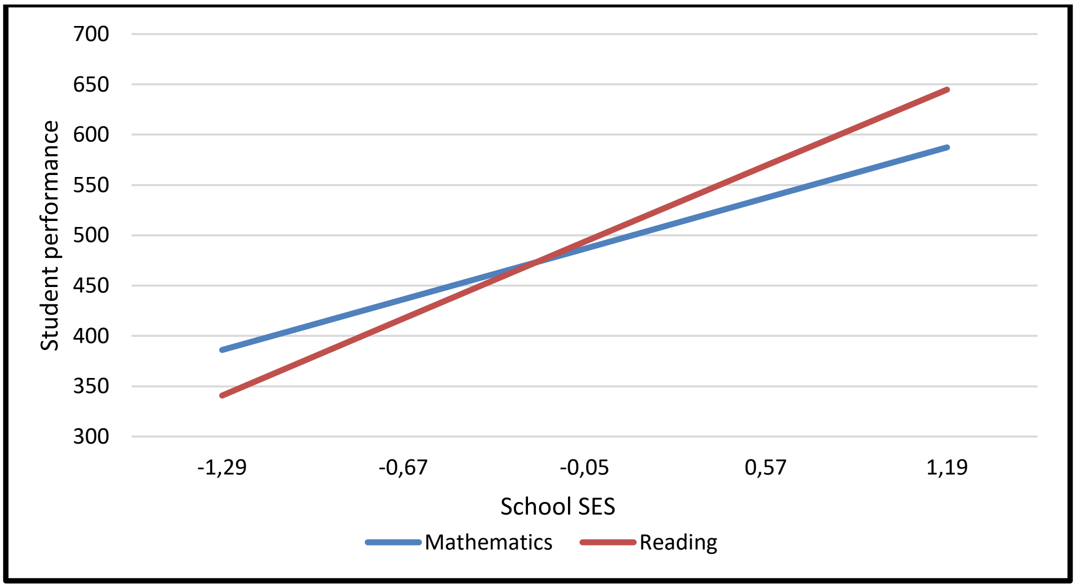
[11] Remember that the main effects of school-level characteristics apply for native students and students with an average SES.

[12] Keep in mind that (1) the effect of the average school SES/% of non-native speakers in a school, may be overestimated due to the measurement error of individual SES, and (2) that our research focuses on systemic effects throughout the school career: it aims to capture the *cumulative* impact of social inequalities on cognitive skills at age 15, whatever their origin (social capital acquired in the students' home environment as well as segregation during the entire past school career) and whatever

can also calculate the achievement gaps in mathematics and reading scores of 15-year-olds between schools with a very low SES (5th percentile) and very high SES (95th percentile) (Table 1). The achievement gap in mathematics (see Figure 5) is 162.57 points, which is equal to more or less 4 years of formal schooling. The achievement gap in reading is even larger: 247.13 points, which is equal to circa 6 years of formal schooling (see Figure 5). Both achievement gaps are considerable, indicating that socioeconomic school segregation is one of the main barriers that should be broken down in order to increase student-equity in education.

Finally, in both models the association between the percentage of non-native speaking students in a school and student performance is negative and significant (Model 1C, 2C), indicating that a higher share of non-native speaking students in a school goes hand in hand with lower performance of all (native speaking) students. Both coefficients are, however, very small: a 20% increase in the non-native share of a school, leads to a decrease by 0.31 and 0.63 points in resp. native students' mathematics and reading scores.

Figure 5. Student mathematics and reading performance by school SES



4.1.4. School-level variables and socioeconomic and language-based student-equity

The cross-level interaction terms between students' SES or language spoken at home with the school-level characteristics enable us to examine if (and how) socioeconomic and/or language-based student-inequity varies between schools. Results in Table 3 (Model 1C, 2C) show that some school-level characteristics are indeed significantly associated with the socioeconomic and/or language-based student-inequity.

Before discussing the results, we remind the reader that the graphs presented in this section are only showing the changes in the SES-gradient and/or language-gradient, without taking into account the

the stage at which they were shaped. Both school composition effects at age 15 also capture the students' shared social capital, which is (at least partly) determined by progressive segregation mechanisms rooted in the education system.

main effects. To get a fuller picture, the main effects must be considered jointly with the interaction effects.

First, the percentage of fully certified teachers is positively correlated ($p<0.001$) with socioeconomic student-inequity in mathematics, and negatively with language-based student-inequity in mathematics. More precisely, for mathematics performance, an increase by ten percentage points of the proportion of fully certified teachers, will increase the socioeconomic student-inequity coefficient by 0.09 points (SES-gradient increases from 16.52 to 16.61), but reduce the performance gap between native and non-native students by 2.46 points (language gradient diminishes from -21.99 to -19.53). In practical terms, the 'increase' in the socioeconomic student-inequity is negligible, and thus we conclude that non-native speaking students in particular benefit from attending schools with a higher percentage of fully certified teachers.

Next, we find small, significant correlations between the student-teacher ratio (at school level) and socioeconomic (Model 1C, B=0.02) and language-based student-equity (Model 1C, B =0.65) in mathematics. Yet, both coefficients are very small suggesting that the student-teacher ratios do not substantially affect students' reading and mathematics performance.

Third, with respect to the correlation between the % of non-native speakers in a school and socioeconomic and language-based student-inequity, we observe in both models a weak (though significant) negative correlation with the language-based student-inequity, but no significant correlation with socioeconomic student-equity. More precisely, an increase by ten percentage points of non-native speaking students in a school, would go hand in hand with a 0.12 / 0.14 points reduction in the student language gap amongst 15-year-olds. Again, this increase is very small, indicating that the percentage of non-native speakers in a school is barely related to the student-inequity in a school.

Finally, we find significant interaction effects between a school's average SES and both the socioeconomic and language-based student-inequity (Model 1C, 2C). The interaction effect between a school's SES and the socioeconomic student-inequity (see Figure 6) is positive and significant in the two models (Model 1C, B = 5.76; Model 2C, B = 2.03), suggesting that schools with a higher SES, have higher levels of socioeconomic student-inequity in mathematics and reading. Figure 6 presents this graphically. In schools with a very low average SES, the impact of student SES on reading performance is close to zero, compared to 8 points per standard deviation of student SES in schools with a high SES. The student-SES gradient for mathematics has a value of 5 points in low SES schools, compared to a value of 28 in schools with a high SES. As regards the interaction between school-SES and the language-based student-inequity, our analyses are rather ambivalent (see Figure 7). In Model 1C, in which mathematics performance is analysed, we observe a positive interaction effect (B=2.59), suggesting that the language-based student-inequity improves in schools with a higher SES. More precisely, in schools with a very high SES (95^{th} percentile), the achievement gap in mathematics between native speaking and non-native speaking students is 10 points lower than in schools with a very low SES (5^{th} percentile). The opposite association is found in Model 2C (B=-7.38), in which reading performance is analysed: the higher the school SES, the larger the achievement gap in reading between native and non-native speaking students (the gap between native and non-native speakers increases by 30 points between the 5th and the 95th percentile of schools, ranked by average SES).[13]

Figure 6. Socioeconomic student-inequity by school SES

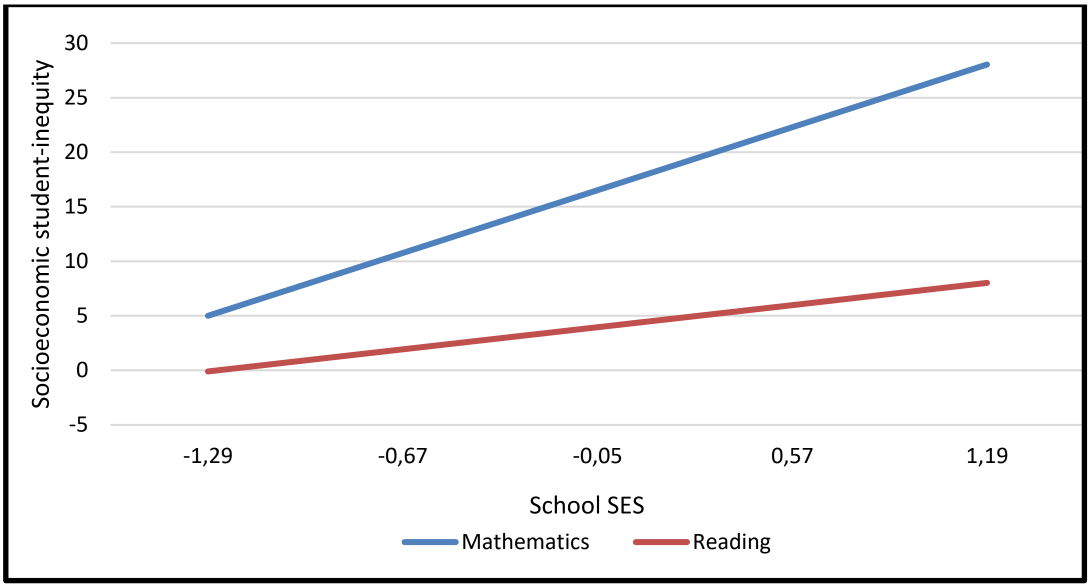
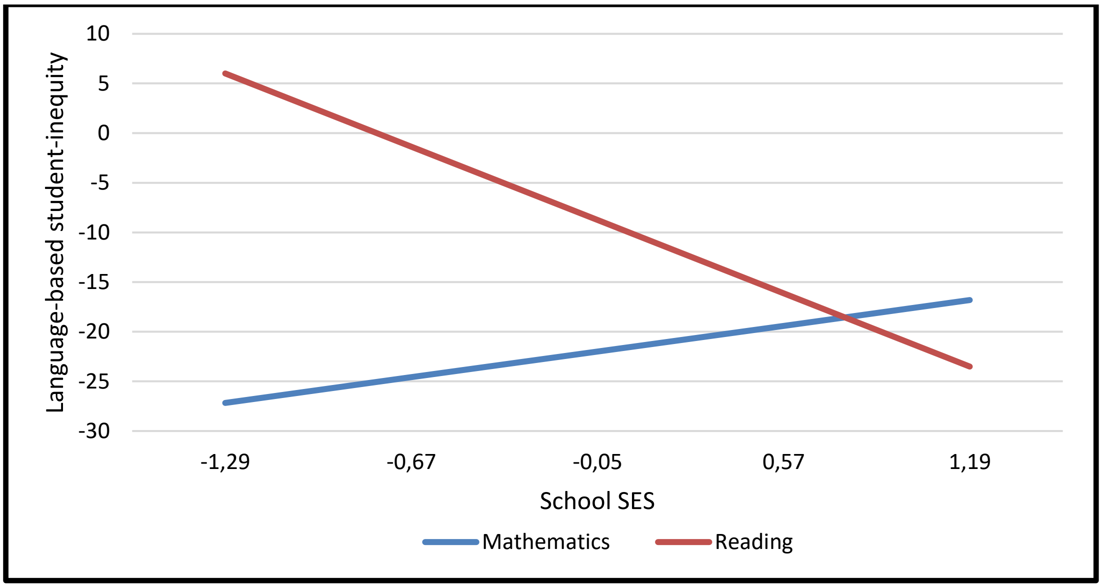


Figure 7. Language-based student-equity by school SES



In sum, two school-level characteristics are not or only very weakly related to student-equity: the student-teacher ratio and the percentage of non-native speaking students in a school. Both school-level factors will, in other words, not affect the level of student-equity in any substantial way. The percentage of fully certified teachers is positively and significantly correlated with the language-based student-equity as it reduces the achievement gap in reading between native and non-native 15-year-

[13] Remember that figures 6 and 7 only reflect *interaction effects* with school-SES, not the overall effect of school-SES. For example, it would be wrong to conclude from Figure 7 that non-native students are performing better in reading in low SES schools than in high SES schools. When combining the main and interaction effects of the school SES, native as well as non-native speaking students perform considerably better in high SES schools.

old students. The school SES is positively correlated with the level of socioeconomic student-equity, meaning that the achievement gap between advantaged and disadvantaged 15-year-old students will be larger in school with a higher SES compared to schools with a lower SES. Or, students with a high SES will benefit most from going to high SES schools. The correlation between school SES and the language-based student-inequity is inconclusive: it is positive for mathematics, but negative for reading performance. Hence, the higher the school SES, the larger the performance gap in reading between native and non-native speaking students, but the smaller the achievement gap in mathematics.

4.2. System-level characteristics, student performance and student-equity

The intercepts in Models 3C and 4C (Table 4) reflect the average score of a non-immigrant, native speaking, female student with an average age (15.79), an average SES (-0.05), that lives in an early-tracking education system (<15) with an average GDP per capita (34,100), an average percentage of students that have repeated a grade (11.12%), an average level of competitiveness between schools (46.69%) and in which the government spends an average amount of money on primary education (21.01).

Similar to the changed interpretation of the main effects in Models 1C and 2C after including interaction terms, the main effects in Models 3C and 4C should also be interpreted differently. In models 3C and 4C, system-level characteristics and student-level characteristics are included as well as cross-level interaction terms between all these characteristics. Consequently, the interpretation of the main effects of the student-level factors is: the correlation between a students' SES/language spoken at home and student performance *in an early-tracking education system (<15) with an average GDP per capita (342,000), an average percentage of students that have repeated a grade (11.12%), an average level of competitiveness between schools (46.69%) and in which the government spends an average amount of money on primary education (21.01).* Similar to the interpretation of the school-level characteristics, the effects of the system-level characteristics only apply *for native speaking students with an average SES*.

4.2.1. Control variables

Results regarding our control variables (age, gender and migration status) are in line with the results of our previous analyses (Model 1C, Model 2C, Table 3). The age of a student shows a significant positive correlation with both mathematics and reading performance of about the same effect-size (Model 3C, B=12.55; Model 4C, B = 12.85). The migration status is in both models negatively correlated with mathematics and reading student performance, indicating that immigrant students perform worse than non-immigrant students. However, this correlation is not significant, which is probably due to the fact that students' SES and students' language spoken at home are also included in the models and already capture the negative effect of being an immigrant student. The correlation of gender with students' mathematics and reading performance shows an opposite sign. Across all countries, boys seem to outperform girls by 7.09 points in mathematics, but girls outperform boys in reading by 27.50 points.

Table 4: System-level characteristics and student performance

	Model 3C: Math		Model 4C: Reading	
	B	**SE**	**B**	**SE**
Intercept	524.28***	7.59	540.19***	8.08
Level 1: student-level				
Age *(centered)*	12.55***	0.0	12.85***	1.08
Migration *(0 = non-immigrant student)*	-3.84	3.91	-1.99	4.00
Language *(0 = native speaker)*	-32.73***	4.13	-37.27***	4.45
Gender *(0 = female)*	7.09***	1.06	-27.50***	1.46
SES *(centered)*	36.14***	1.64	37.46***	1.70
Level 3: system-level				
Competition between schools (in %)	1.46**	0.50	1.74**	0.60
GDP per capita/10000	8.98***	2.58	10.48***	2.84
% students that repeated a grade	-0.79	0.61	-0.89	0.73
Tracking age (0 = <15)	-21.88†	12.00	-6.14	11.82
Expenditure on primary education (per student)	0.44	1.37	0.44	1.82
Random slope system-level				
SES slope	55.11***	13.37	66.04***	16.88
Language slope	31.22***	81.19	315.45***	74.04
Cross-level interactions				
Expenditure on primary education*SES	-0.06	0.27	0.01	0.32
GDP per capita/10000*SES	0.33	0.41	0.42	0.47
% grade repetition*SES	-0.26*	0.11	-0.27*	0.11
Tracking age*SES	-3.84†	2.23	-4.73*	2.39
% competition*SES	0.16†	0.09	0.15	0.11
Expenditure on primary education*lang	0.37	0.56	0.41	0.69
GDP per capita/10000*lang	0.84	0.75	-0.95	0.92
% grade repetition*lang	0.42*	0.18	0.47*	0.21
Tracking age*lang	9.57†	5.56	4.37	5.28
% competition*lang	-0.05	0.22	-0.22	0.23

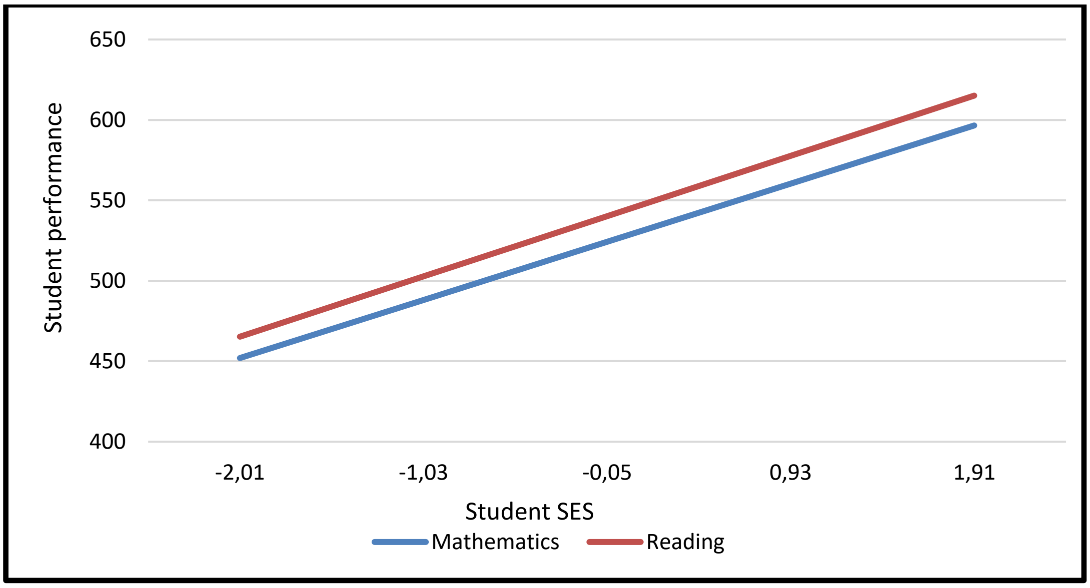
***p<0.001, **p<0.01, *p<0.05, † p<0.1

4.2.2. Student-level characteristics and student performance

Similar to the results of our previous analyses (Model 1C, Model 2C), we observe a negative significant correlation between language spoken at home and student performance in both models, meaning that – across all countries – 15-year-old non-native speakers perform worse than native speakers (Model 3C, B=-32.73; Model 4C, B=-37.27). The correlations between language spoken at home and student performance are clearly higher than in our previous models (Table 3, model 1C, 2C). This is due to the fact that we exclude school-level predictors in this model and consequently, the correlations with individual student characteristics are partly over-estimated as they also capture the effect of language-based school segregation.

Concerning the second student-level characteristic – SES – a positive significant correlation is found with both mathematics and reading student performance. This indicates that – across all countries – students with a higher SES outperform students with a lower SES. More specifically, students with a higher SES (1 standard deviation above the average), score 36.14 points higher on mathematics and 37.46 points higher on reading than students with an average SES. Based on the descriptive table (Table 1), we can estimate the difference in mathematics and reading performance between low-SES students (5th percentile) and high-SES students (95th percentile). The achievement gap in reading and mathematics is more or less the same: resp. 117.25 and 113.12 points (see Figure 8). This gap is almost equal to three years of formal schooling, but it should be taken into account that the correlation between students' SES and student performance is partly over-estimated as it also captures the effect of socioeconomic school segregation due to the fact that we excluded school-level predictors in these models.

Figure 8. Student mathematics and reading performance by students' SES



4.2.3. System-level characteristics and student performance

Looking at Table 4, we observe that students' mathematics and reading performance are positively correlated with the GDP per capita of their country (resp. B=8.98, B=10.48) and the level of competitiveness between schools (resp. B=1.46, B=1.74). More specifically, in countries where the GDP per capita increases by $10,000/year, 15-year-old students will score 4.12 points higher in mathematics and 4.81 points higher in reading, than students living in countries with an average GDP per capita (34,100). Similarly, in countries where the competition between schools increases by 10 percentage points, students will perform 1.26 points better in mathematics and 1.50 points better in reading, than in countries with an average level of competitiveness. Although both coefficients are significant, it seems that their magnitude is rather small: student performance changes little depending on a country's GDP per capita or the level of competitiveness of its education system.

Another system-level characteristic that shows a significant correlation with student performance – although only with student mathematics performance – is tracking age. Table 4 (Model 3C) reveals a negative, quite large, correlation between the tracking age and student mathematics performance (B=21.88). This indicates that students in early tracking countries (<15), score 21.88 points higher for mathematics than students in late-tracking countries (≥15). However, the coefficient is only significant at 0.1% level.

The last two system-level characteristics – frequency of grade repetition and government expenditure on primary education – are not significantly correlated with student performance. This indicates that across all countries, neither the incidence of grade repetition nor government expenditure on primary education matter in terms of reading and mathematics performance of 15-year-old students.

4.2.4. System-level variables and socioeconomic and language-based student-inequity

Again, the cross-level interaction terms between students' SES or language spoken at home with the system-level characteristics enables us to examine if / how socioeconomic and/or language-based equity vary with system-level characteristics.

Similarly to what we mentioned in the previous section (4.1.4), graphs presented in this section only show changes in the SES-gradient and/or language-gradient without taking into account the main effects. These effects must, however, be taken into consideration to get the full picture.

First, the frequency of grade repetition in a system is significantly correlated with the socioeconomic and language-based student-inequity (Table 4, Model 3C and 4C). With respect to the student-level socioeconomic gradient, we observe a negative correlation (Model 3C, B=-0.26; Model 4C, B=-0.27) in both models, indicating that the socioeconomic student-inequity is lower in education systems with higher levels of students that have repeated a grade (see Figure 9). However, the correlation is negligible: an increase by 10 percentage points of students that have repeated a grade, goes in pair with a decline of approx. 0.21 in the socioeconomic student-inequity in the two models. In other words, the SES-gradient remains quite high (Model 3C, B=35.83; Model 4C, B=37.252). As for language-based student-inequity, a significant positive coefficient is found (Model 3C, B=0.42; Model 4C, B=0.47) in both models, indicating that education systems with higher frequencies of grade repetition are characterised by lower levels of language-based student-inequity (see Figure 10). Again, the coefficient is very small and is almost negligible. Based on previous research, these results were not expected. Most studies that have focussed on the relationship between the percentage of students that have repeated a grade in an education system and student performance found that grade repetition is associated with higher levels of student-inequity as the achievement gap between disadvantaged (in terms of SES and language) and advantaged students, enlarges. However, they also stated that– apart from the negative correlation with student performance – grade repetition has negative financial and social effects. For example, grade repetition is a source of stress, ridicule and bullying by others (Yamamoto and Byrnes, 1987; Anderson, Jimerson and Whipple, 2005), which negatively affects self-esteem and increases the likelihood of high-risk behaviour, school failure and dropout.

Figure 9. Socioeconomic student-inequity by % of students that have repeated a grade

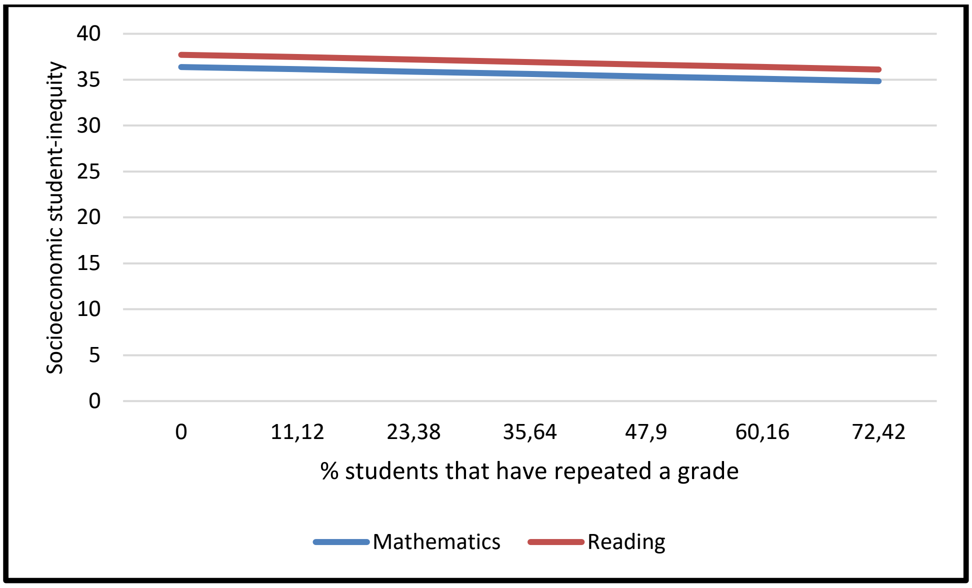
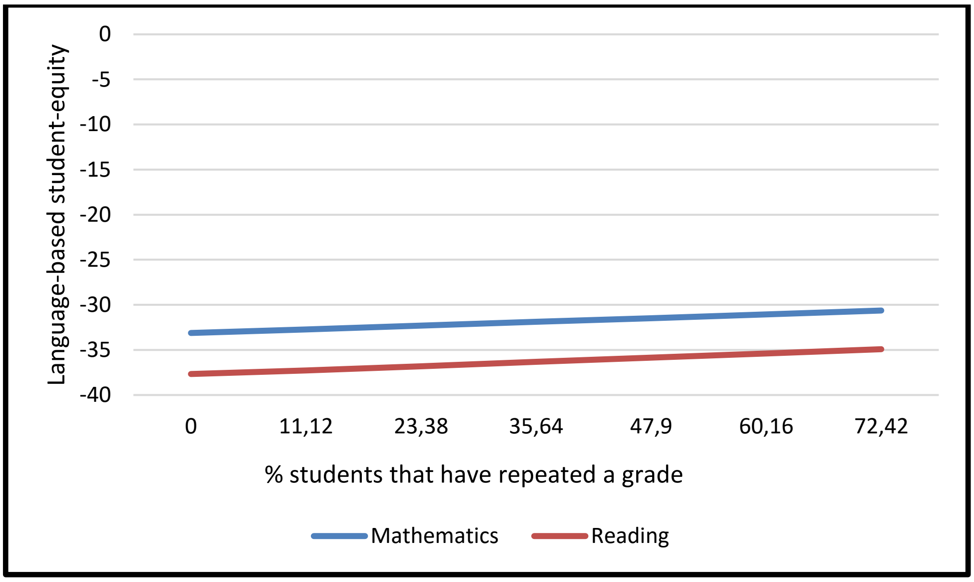


Figure 10. Language-based student-inequity by % of students that have repeated a grade



Secondly, a significant negative correlation between the age that students are tracked in an education system and socioeconomic student-inequity is found in both models (Model 3C B=-3.84; Model 4C B=-4.73), indicating that – ceteris paribus - late-tracking education systems (≥15) have higher levels of socioeconomic student-equity in reading (the SES-gradient shrinks from 36.14 to 32.73) and mathematics (the SES-gradient reduces from 37.46 to 32.73) than early-tracking education systems (<15) (see Figure 11 and 12). The correlation between the tracking age of students and language-based student-inequity is less clear. In the model where mathematics performance are modelled (Model 3C), we find a positive, fairly large, coefficient (B=9.57) that is significant at p<0.1 level. This indicates that late-tracking education systems have higher levels of language-based student-equity as the achievement gap in mathematics between native and non-native students is smaller. We do not find any significant correlation in the model where language-based student-inequity in reading performance are analysed (Model 4C).

Figure 11. Socioeconomic student-inequity by tracking age

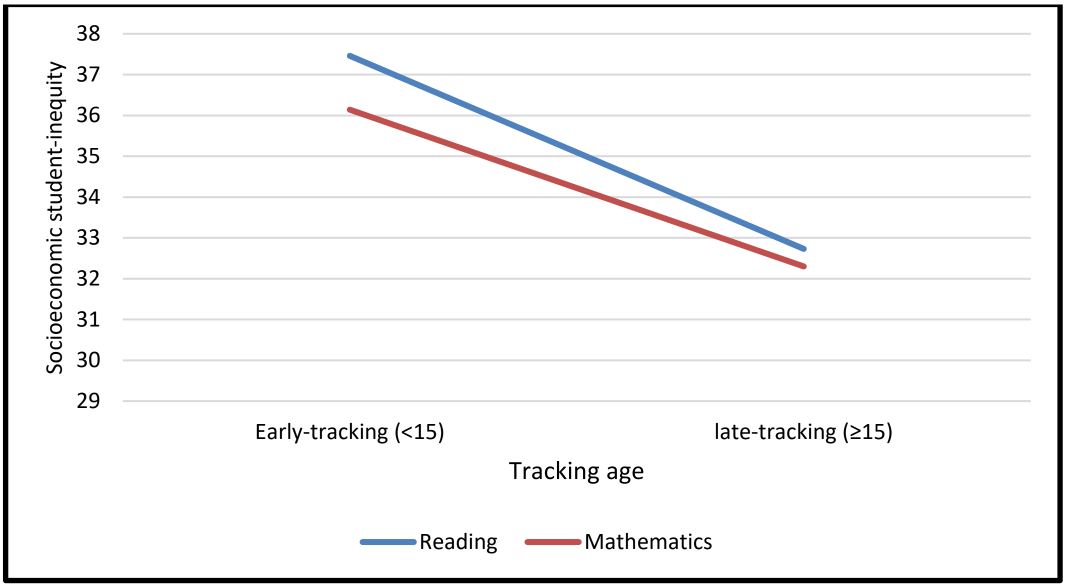
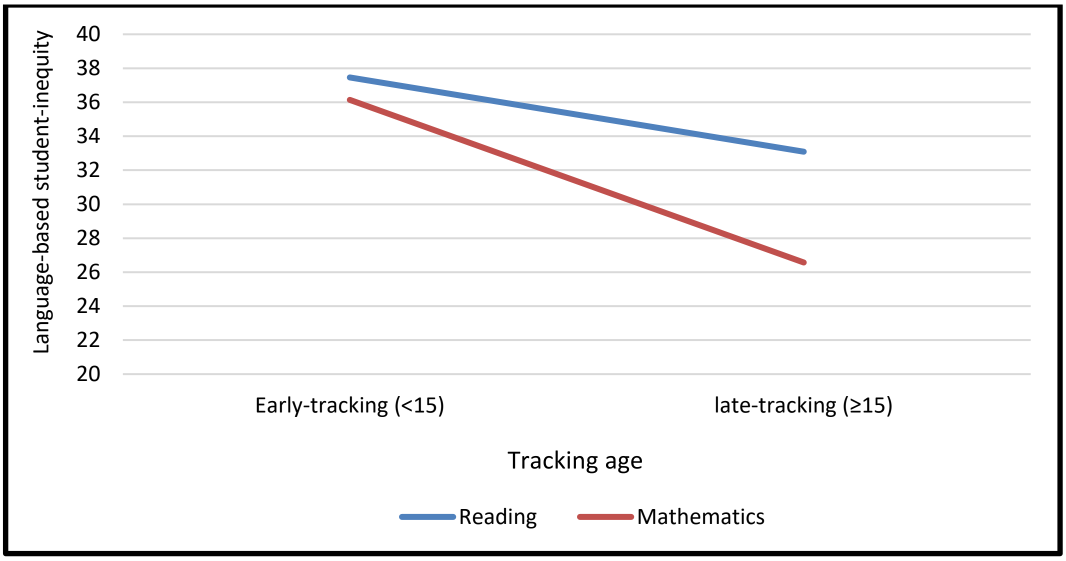


Figure 12. Language-based student-inequity by tracking age



Third, we observe a very weak, significant ($p<0.1$), positive correlation between the level of competitiveness between schools and the socioeconomic student-inequity in mathematics (Model 3C, B=0.16). Given that we do not find any other significant correlation between the competitiveness between schools and student-inequity, we conclude that the level of competitiveness (as measured in our model) is not related to the level of student-inequity.

The other system-level characteristics do not show any significant correlation with neither socioeconomic student-inequity nor language-based student-inequity, meaning that the level of student-inequity does not vary according to an education system's level of expenditure on primary education and its GDP per capita.

In sum, the tracking age of students is the only system-level characteristics that is clearly related to the socioeconomic and language-based student-inequity. In particular, our analyses reveal that late-tracking education systems (≥15) have substantially lower levels of socioeconomic and language-based student-inequity than early-tracking education systems (<15), indicating that disadvantaged students or students with another mother tongue profit of being tracked at later ages. The other system-level characteristics only showed weak correlation with socioeconomic and/or language-based student-inequity, or none at all.

4.3. System-level characteristics, school performance and school-equity

In our last set of analyses, we examined the relation between system-level characteristics on the one hand, and school performance and school-inequity on the other. The intercepts in both models (Table 5, Models 5C and 6C) reflect the average score of a school with an average student-teacher ratio (12.84), an average school-SES (-0.05), an average percentage of non-native speaking students (12.86%) and an average percentage of fully certified teachers (87%), located in an early-tracking education system (<15) with an average GDP per capita (34,100), an average percentage of students that have repeated a grade (11.12%), an average level of competitiveness between schools (46.69%) and in which the government spends an average amount of money on primary education (21.01).

Table 5. System-level characteristics and school performance

	Model 5C: Math		Model 6C: Lang	
	B	**SE**	**B**	**SE**
Intercept	471.31***		463.86***	
Level 2: school-level				
% certified teachers *(centered)*	6.10*	2.94	8.04***	3.15
Student-teacher ratio *(centered)*	0.59***	0.14	0.55***	0.15
School SES *(centered)*	79.02***	7.04	81.50***	5.85
% non-native speakers in a school *(centered)*	-0.61***	0.13	-0.66***	0.13
Level 3: system-level				
Competition between schools (in %) *(centered)*	1.09*	0.43	1.25***	0.39
GDP per capita/10000 *(centered)*	9.56***	2.40	9.20***	2.29
% students that have repeated a grade *(centered)*	-0.15	0.37	-0.20	0.42
Tracking age *(0 = <15)*	-4.00	9.77	5.30	9.22
Expenditure on primary education *(centered)*	0.87	1.03	1.06	1.04
Random slopes system-level				
School SES	650.79***	289.61	600.94***	146.15
% non-native speakers	0.42†	0.24	0.31***	0.16
Cross-level interactions				
Expenditure on education*school SES	0.28	0.81	0.11	0.73
% competition*school SES	0.65†	0.38	0.67†	0.36
% grade repetition*school SES	-0.23	0.31	-0.34	0.32
Tracking age*school SES	-12.52	8.15	-11.76	7.51
GDP per capita/10000*school SES	1.09	1.30	1.64	1.33
Expenditure on education *%non-native speakers	0.02	0.02	0.01	0.02
% competition*% non-native speakers	0.02†	0.01	0.01	0.01
% grade repetition*% non-native speakers	0.00	0.01	0.00	0.01
Tracking age*% non-native speakers	0.33	0.20	0.23	0.18
GDP per capita/10000*% non-native speakers	0.01	0.04	-0.03	0.04

***p<0.001, **p<0.01, *p<0.05, † p<0.1

Again, we would like to remind the reader of the changed interpretation of the main effects in Models 5C and 6C as we included interaction terms between all system-level characteristics and *school-level* characteristics. The main effects of all school- and system-level variables should be interpreted in the same way as the intercepts. For instance, the effect of the student-teacher ratio on school performance should be interpreted as the correlation in a school an average school SES (-0.05), an average percentage of non-native speaking students (12.86%) and an average percentage of fully certified teachers (87%), located in an early-tracking education system (<15) with an average GDP per capita (34,100), an average percentage of students that have repeated a grade (11.12%), an average level of competitiveness between schools (46.69%) and in which the government spends an average amount of money on primary education (21.01).

4.3.1. School-level characteristics and school performance.

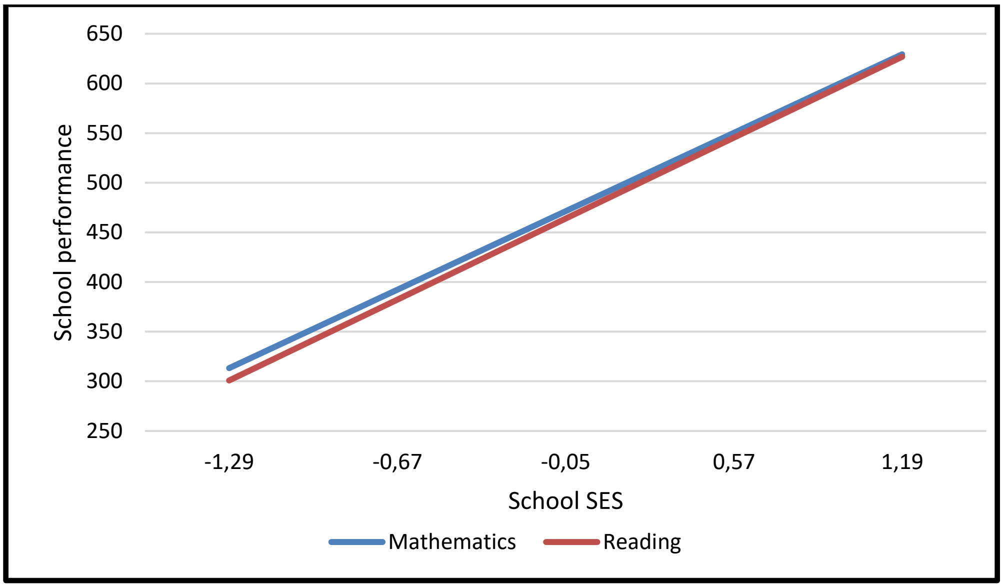
Table 5 (Models 5C and 6C) reveals that all school-level characteristics have a significant correlation with schools' mathematics and reading performance.

First, in both models (Model 5C, B=6.10; Model 6C, B=8.04) the coefficient of the percentage of fully certified teachers is significant and positive. An 10 percentage points increase in fully certified teachers in a school is associated with an increase of 2.26 and 2.98 points, respectively, in the average mathematics and reading performance of 15-year-olds in a school. This parameter is rather small, indicating that the percentage of fully certified teachers does not impact schools' average performance considerably.

Secondly, the student-teacher ratio is positively correlated with school performance and has more or less the same effect-size in both models (Model 5C, B=0.55; Model 6C, B =0.59). Hence, the higher the student-teacher ratio, the higher the average mathematics and reading performance in a school. The effect size is, however, negligible: 5.42 students more per class in a school goes hand in hand with an increase of 0.55 and 0.59 points in schools' average mathematics and reading performance. This means that the student-teacher ratio is not substantially related to school performance.

Thirdly, a large and significant positive correlation between school SES and schools' mathematics (B= 79.02) and reading performance (B= 81.50) is revealed (Model 5C, 6C). This indicates that the school SES is a strong determinant of schools' cognitive outcomes. Similarly to our calculations in the previous models, we can estimate the achievement gap in schools' average mathematics and reading performance between schools with a very low (5^{th} percentile) and very high (95^{th} percentile) school-SES. The achievement gaps are enormous (see Figure 13): 254.9 points in schools' mathematics performance and 262.9 points in schools' reading performance. Remember that our research focuses on systemic effects throughout the school career: the school-SES effect captures the cumulative impact of social inequalities on cognitive skills at age 15, whatever their origin (social capital acquired in the students' home environment as well as segregation during the entire past school career) and whatever the stage at which they were shaped.

Figure 13. School mathematics and reading performance by school SES



Finally, the association between the percentage of non-native speaking students in a school and school performance is negative and significant in both models (Model 5C, 6C), indicating that a higher share of non-native speaking students in a school goes hand in hand with lower school performance. Both correlations are, however, again very small: an increase of 20 percentage points in the non-native share of a school, leads to a decrease of 0.60 and 0.65 points in resp. schools' average mathematics and reading scores.

4.3.2. System-level characteristics and school performance

Table 5 reveals that two system-level characteristics are significantly associated with school performance in mathematics and reading: GDP per capita (Model 5C, B=9.56; Model 6C, B = 9.20) and the level of competitiveness between schools (Model 5C, B=1.09; Model 6C, B = 1,25). Both system-level characteristics show a positive correlation with schools' mathematics and reading performance. More specifically, in countries where the GDP per capita increases by $10,000/year, the mathematics and reading performance of a school increases by resp. 4.39 and 4.22 points. Similarly, in countries where the competition between schools increases 10 percentage points, schools' mathematics and readings performance will increase by 0.94 and 1.08 points. Although both system-level characteristics show a significant correlation with school performance, their magnitude of their effects is very small.

The other system-level characteristics – the tracking age, frequency of grade repetition and government expenditure on primary education – are not significantly correlated with average school performance. This indicates that across all countries, the tracking age, the percentage of students that have repeated a grade and the government expenditure on primary education does not matter in terms of schools' average reading and mathematics performance.

4.3.3. System-level variables and socioeconomic and language-based school-inequity

Cross-level interaction terms between a schools' SES or the % of non-native speakers in a school with system-level characteristics, enable us to examine if / how socioeconomic and/or language-based school-inequity vary with system-level characteristics. However, results in Table 5 suggest that most of our system-level characteristics do not impact the school-inequity substantially. We only observe a few small correlations that are significant at the 10%-level. More specifically, we find a weakly significant ($p<0.1$) positive correlation between an education system's level of competitiveness between schools and the socioeconomic school-inequity in mathematics (Model 5C, B=0.65) and in reading (Model 6C, B=0.61). An increase of 10 percentage points in the level of competitiveness is associated with an increase of resp. 0.56 and 0.53 in the socioeconomic school-inequity. Furthermore, we observe a significant correlation between the level of competitiveness and the language-based school-inequity in schools. However, this coefficient is almost equal to zero, so we can conclude that language-based school-equity is not related to the level of competitiveness in an education system.

None of the other system-level characteristics show any significant correlation with socioeconomic school-inequity, nor with language-based school-inequity. This suggests that other system-level factors might be better suited in explaining differences in school-inequity.

5. Discussion and implications

The aim of this report was to analyse how student (RQ1) and school (RQ3) performance, as well as student- (RQ2) and school-inequity (RQ4) are correlated with school- and system-characteristics. We proposed four research questions (RQ) to answer this general question:

RQ1	Which school- and system-level characteristics are correlated with the *mathematics and reading performance* of 15-year-old students and how?
RQ2	Which school- and system-level characteristics are correlated with *socioeconomic and language-based student-inequity* and to what degree does socioeconomic and language-based student-inequity vary with these characteristics?
RQ3	Which school- and system-level characteristics are correlated with the average mathematics and reading performance of all 15-year-old students *at school level* and how?
RQ4	Which system-level characteristics are correlated with *socioeconomic and language-based school-inequity* and how?

In this section, we will formulate an answer to each research question separately (apart from RQ1 and RQ3) based on the results in section 4. Research questions 1 and 3 will be discussed simultaneously as school performance is measured by the average performance of 15-year-olds in a school. Consequently, we assume that results are more or less the same for both student and school performance.

5.1. How are school- and system-level characteristics correlated with *mathematics and reading student and school performance*?

Models 1C, 2C, 3C, and 4C (Table 3 and 4) were estimated to answer RQ1 (student performance), and Models 5C, 6C (Table 5) were estimated to answer RQ3 (school performance). In what follows, we will first focus on school-level characteristics (Table 3 & 5) and afterwards on system-level characteristics (Table 4 & 5).

5.1.1. School-level characteristics.

We included four school-system level characteristics: (1) the percentage of fully certified teachers – as a proxy for teacher quality, (2) the percentage of non-native speaking students in a school, (3) the student-teacher ratio and (4) the (cumulative) school SES – as a proxy for the cumulative socioeconomic school segregation.

In general, we found that the percentage of fully certified teachers and the (cumulative) school SES are positively associated with student and school performance. The percentage of non-native speaking students is negatively correlated with student and school performance. The student-teacher ratio shows a positive correlation with school performance, but a negative correlation with student performance (Table 3 & 5, Models 1C, 2C, 5C, 6C).

With respect to teacher quality, our analyses showed a positive correlation with student and school performance. This is in line with most of the existing literature (Clotfelter, et al., 2007; Clotfelter et al., 2010; Darling-Hammond, 2004; Monk, 1994; Ronfeldt et al., 2012; Caena, 2011; OECD, 2005): the higher the quality of teachers, the better the student and school performance. Ensuring a sufficient supply of qualified teachers thus remains an important responsibility for governments, all the more because there is a risk that shortages in the teacher market induce Matthew effects: very often, disadvantaged schools are the first victims of shortages of qualified teachers, as they are less attractive for teachers in general (OECD, 2013a, 2017c).

Our most important finding is that school-SES has a large positive correlation with student and school performance[14]. This suggests that the socioeconomic school segregation is a strong determinant - stronger indeed than students' individual SES - of inequities in cognitive outcomes between students and schools: students attending an advantaged school will – regardless of their own socioeconomic background – substantially outperform students attending an average or a disadvantaged school. Conversely, attending a low-SES school results in low achievement for all students (even those from high socioeconomic background). This is in agreement with a large body of evidence – both in primary and secondary education (Van Erwijk et al., 2010, Cumberworth et al., 2018; Agirdag et al., 2012; OECD 2018; Marks, 2015). Several mechanisms are frequently mentioned as possible explanations of the positive coefficient (OECD, 2016A; OECD, 2018; OECD, 2019; Lavy et al., 2012; Burke et al., 2013; Franck et al, 2019):

1. *Peer group effects*: students learn from each other. Given that disadvantaged students often perform less well than advantaged students, it is unfavourable for disadvantaged students to be concentrated in a school as they will learn less from each other and therefore (often) perform worse (Thrupp et al., 2002; Van Eewijck & Slegers, 2010).
2. The experience of '*demotion'* into lower-ability classes, tracks or schools by itself tends to demotivate students and to reduce their ambition and effort (Pelleriaux, 2001; Van Houtte & Stevens, 2010; Spruyt, 2014).
3. Teachers have a tendency to adapt their expectations and standards to the weaker performing students, which may result in downward spirals (the so-called Pygmalion effects - Rosenthal & Jacobson, 1968; Speybroeck, 2013).
4. *Differences in school climate*: disadvantaged schools often have less orderly learning environments. Moreover, teachers frequently lower their expectations in disadvantaged schools and need to spend more time organising and improving the learning environment at the expense of instruction time (Thapa et al., 2013; Gustafsson et al., 2018).
5. *Inequalities in the social and cultural capital of schools*: disadvantaged schools have weaker parental support and can rely less on the social and cultural resources of the parents and the local community as they are often disadvantaged themselves (Poesen-Vandeputte & Nicaise, 2015).
6. *Matthew-effects in the human and educational resources of schools*: although we controlled for the quality and quantity of teachers in our analyses, studies also repeatedly revealed that disadvantaged schools more often face a lack of well-qualified management staff and a lack of available educational resources. This could impact student performance negatively, in

[14] The school SES effect should be interpreted as the proxy for the students' shared social capital at the age of 15, which is (at least partly) determined by progressive segregation mechanisms rooted in the education system

particular in disadvantaged schools (Thrupp & Lupton, 2006; OECD, 2018b; Poesen-Vandeputte & Nicaise, 2015).

7. *Differences in curricula*: in many education systems, students are tracked into different tracks. Disadvantaged students end up more often in 'lower' (less demanding and less valued) tracks – which are more often taught in disadvantaged schools – than advantaged students. Many studies have already confirmed that – ceteris paribus – students allocated to higher tracks outperform identical students in lower tracks (OECD, 2016a; Lavrijsen & Nicaise, 2015; Dockx, 2019). Similar effects occur in comprehensive systems with ability grouping in separate classes.

In sum, we conclude that high levels of socioeconomic school segregation may to some extent be profitable for students attending advantaged schools, whereas they are definitely detrimental for students attending disadvantaged schools. A better social mix within schools could improve in particular the performance of disadvantaged students (as they often attend disadvantaged schools). Moreover, our findings also suggest that promoting socially mixed schools will not lower the average performance level of education systems while it could lower student-inequities significantly.

Third, we find a very small negative association between the percentage of non-native students at school on the one hand, and student and school performance on the other. This coefficient fulfils the expectation we had based on the few studies that already existed: Tonello (2016) found a weak negative correlation between the share of non-native speakers in junior high schools and native peers' language test scores in Italy. Gaey et al. (2013) concluded that an increase in non-native speaking students in primary education does not impact the cognitive outcomes of native speaking students in England. Moreover, she shows that the frequently observed negative correlation in England is rather an artefact of selection: non-native speakers are more likely to attend schools with disadvantaged native speakers. Given that we only find very small correlations, we argue that the percentage of non-native speakers at school barely impacts on student and school performance. Note that we distinguish between migration status, SES and mother tongue in our model: consequently, part of the effect of the percentage of non-native students and students with a migration background at schools is, as proposed by Gaey et al. (2013), captured in the coefficient of the school SES. Immigrants and non-native students attend more frequently schools that are more disadvantaged in terms of SES (OECD, 2016b). This again shows the necessity of striving towards more socially mixed schools.

Finally, the student-teacher ratio has a small negative association with student performance, but a small positive correlation with school performance. Both correlations are negligible and thus we can assume that, in our study, the student-teacher ratio is not related to students' and schools' cognitive outcomes. As no conclusive evidence was found in previous studies, our results are not surprising (Krueger et al., 2001; Ding et al., 2010; Shin, 2012; Bouguen et al., 2017; Bjorklund et al., 2005; Andersson, 2007; Nusche, 2009; Leuven et al., 2008; Li et al., 2016; Shen et al., 2017; Escalle et al.; Jahanshahi et al., 2017). This finding is particularly relevant for the discussion of equity funding – i.e. school funding mechanisms that provide additional resources (including teaching staff) to schools that serve a disadvantaged student population. It might help explain why the effectiveness of equity funding often remains below expectations. Note however that our finding relates to the student-teacher ratio across all schools. More detailed studies focusing on disadvantaged schools, and/or distinguishing between categories of teaching staff may yield more qualified insights.

5.1.2. System-level characteristics

Five system-level characteristics were included in our research: (1) GDP per capita as a proxy for the economic wealth, (2) government expenditure on primary education, (3) the level of competitiveness between schools, (4) the frequency of grade repetition, and (5) the tracking age. The first two characteristics refer to the economic boundary conditions, while the three last variables reflect key institutional features of the education system.

In our analyses (Table 4 & 5, Models 3C, 4C, 5C, 6C), GDP per capita is positively related to student performance, but not to school performance. Remarkably, government expenditure on primary education appears to have no impact: neither on average performance, nor on equity in outcomes. On the other hand, student and school performance are positively associated with the level of competitiveness within an education system. The tracking age is negatively correlated with average student performance in mathematics, but not with student reading performance nor with school performance. Finally, student and school performance are not associated with the frequency of grade repetition.

As regards the relationship between competitiveness and student performance, we found a small positive correlation, which is conform to previous studies (Epple et al., 2017; Hoxby, 2002). Higher levels of competition between schools give schools the 'right' incentives to improve and innovate their education as they are competing for students. Consequently, parents will have a greater choice of schools and can chose the best fitting school in accordance with their educational preferences, as well as the best quality (OECD, 2019; Montes, 2012; OECD, 2017E; Betts and Loveless, 2005; Hoxby, 2000). Our positive coefficients are, however, rather small, indicating that increasing competition between schools is not a very effective tool to increase student and school performance.

Our results regarding the relationship between GDP per capita and mathematics and reading performance are illuminating. Few studies have examined this association, and the ones that did focus on this relationship did not find any relationship or only a weak positive one (Perry, 2009; Pfeffer, 2009; Hout and Diprete, 2006). Our results show that education systems in more wealthy countries have better student performance. As mentioned in our survey of the literature (section 2.3.1), this positive correlation could be interpreted in several ways. On the one hand, more economic wealth in a country raises the demand for a higher-educated workforce. Consequently, parents will be more optimistic about the expected returns to education for their children, and therefore support and motivate their children better during their school career. This could lead to higher student performance. Moreover, irrespective of the country's expenditure on formal education, a more wealthy environment facilitates informal learning (outside school) through cultural and recreational activities, and reduces the opportunity cost of learning. On the other hand, in the long run, higher educational performance in turn leads to increased human capital, which eventually results in more economic growth (Gylfason, 2003; Woessmann, 2008; OECD, 2012c). Either way, we can conclude that in countries with greater economic wealth, 15-year-olds perform better in education. Hence, when comparing education systems in terms of performance and equity levels, the economic wealth should always be taken into account.

Furthermore, we find a negative effect of tracking age on student performance in mathematics, meaning that in late-tracking education systems, students perform less well on average. However, the coefficient was significant only at 10%-level in model 1C. In all other models – 15-year-olds reading performance, school-average mathematics performance and school-average reading performance – the corresponding coefficients were not significant. Hence, the negative effect observed in Model 1C does not seem to be robust. This can be explained by various factors. Firstly, we included some non-OECD countries where tracking is non-existent while the curriculum is rather weak across the board, which could possibly explain why we find a negative effect of late-tracking on mathematics performance. Moreover, it should also be noted that we did not control for initial performance levels of students. Lavrijsen & Nicaise (2015, 2016) did control for performance in grade 4 of primary school and found large positive effects of later tracking for the weakest groups of students, while the top performers did equally well under early and late tracking. In sum, based on those studies, we could assume that late tracking produces gains on average, combined with more equitable outcomes. Whereas this applies to maths and science as well as reading, the effects on reading are the most positive.

Finally, student performance appears to be unrelated with the incidence of grade repetition in an education system nor with government expenditure on primary education. With respect to the practice of grade repetition, our results do not confirm previous studies that found a negative relation (OECD, 2012d; Brophy, 2006). Apart from this negative effect on student performance, these studies also showed that the practice of grade repetition has long-term negative financial, social and academic effects on students. Our findings therefore do not really contradict the earlier studies, given that with PISA data, we are able to measure only student performance halfway in secondary education. It remains plausible that grade repetition does affect other student outcomes such as lower self-esteem, increased high-risk behaviour, higher risk at school failure and dropout, lower earnings at the labour market, etc. (Jacob and Lefgren, 2009; Yamamoto and Byrnes, 1987; Anderson, Jimerson and Whipple, 2005).

The absence of a significant relationship between student performance and government expenditure on primary education, is less surprising as existing results are inconclusive. Some studies did find a positive coefficient, others did not. The OECD (2012c) stated that in countries where the cumulative public expenditure per student from age 6 to age 15 exceeds 35,000 USD, the significant positive relationship with student performance disappears. This possibly explains the absence of a significant correlation in our estimations, as we only included developed countries, which often have quite large government expenditures on primary education.

5.2. Which school- and system-level characteristics are correlated with *socioeconomic and language-based student-inequity* and to what extent does socioeconomic and language-based student-inequity vary with these characteristics?

In order to answer this research question, we included cross-level interaction terms in Models 1C, 2C, 3C and 4C between our two measures of student-inequity – the SES-gradient and the language-gradient – with all school- and system-level characteristics. Similarly to the previous section (5.1.), we will first discuss school-level characteristics (Table 3, Models 1C, 2C) and next system-level characteristics (Table 4, Models 1C, 2C).

5.2.1. School-level characteristics

Overall, we observed that two school-level characteristics – the student-teacher ratio and the percentage of non-native speaking students in a school – are not or only very weakly related to both the socioeconomic and/or language-based student-inequity. The percentage of fully certified teachers is negatively correlated with the language-based student-inequity in mathematics, but not in reading; its correlation with the socioeconomic student-inequity is positive but negligible. School SES interacts positively with socioeconomic student-inequity. Its association with the language-based student-inequity is ambivalent: we found a positive coefficient for reading performance, but a negative coefficient for mathematics performance.

To begin with, the negative effect of teacher quality on language-based student-inequity in mathematics means that non-native speaking students in particular benefit from attending schools with more qualified teachers. Taking into account that more qualified teachers in a school also improve student performance (see section 5.1.1.), we conclude that investing in qualified teachers is a win-win situation for students: it both improves student performance and diminishes language-based student-inequity.

Second, we observe a positive interaction between the (cumulative) socioeconomic school segregation and the socioeconomic student-inequity. This indicates that advantaged schools are confronted with higher levels of socioeconomic student-inequity than disadvantaged schools, or in other words, advantaged students are more sensitive to the school SES than disadvantaged students. Although the opposite effect is often assumed, our analyses and the few previous studies that do exist, suggest otherwise (Palardy, 2008; Wells, 2010; Cumberworth, 2018). Not much research has yet examined the mechanisms that could explain this positive interaction. We suspect two possible explanations:

1. *Stigmatisation*: low-SES students may be more stigmatised or discriminated in high-SES schools. This may lead to a lower level of motivation, a lower self-esteem, a lower well-being at school, etc. amongst disadvantaged students in high SES schools, which often negatively affects their performance.
2. *The social and cultural capital of students:* teachers' expectations are in general higher in advantaged schools and consequently they expect students to learn more easily. Students from privileged backgrounds can rely on the cultural and social capital of their family and local cummunity, which is not the case for disadvantaged students.

Whatever the explanation may be, our estimations point at a major risk that school segregation reinforces socioeconomic inequalities in performance. The lesson for policy making is that segregation should be minimised. This may entail direct measures (such as government intervention in the allocation of students to schools) as well as indirect measures (such as later tracking). Whenever such policies are not accepted or do not work, this would justify a 'non-linear' equity funding mechanism, with more than proportionate extra-support to low-SES schools with a concentration of low-SES students, on top of the lump-sum supplement per disadvantaged student.

Note that the argument relating to SES does not apply to students with a different mother tongue. The interaction between the school SES and the language-based student-inequity is ambivalent indeed. In mathematics, we observed a positive interaction effect between school-SES and the language-related performance gap: the higher the school SES, the lower the language-based student-inequity. By contrast, we found a negative correlation in reading, implying that the language-based student-inequity is larger in high SES schools than low SES schools.

The two remaining school-level factors – student-teacher ratio and the % of non-native speaking students in a school – are not or only barely correlated with student-inequity. These results are in line with our results regarding student and school performance: both effects were negligible (see section 5.1.1.). A possible explanation for the absence of any 'significant' effect of student-teacher ratio could be that we only looked at 15-year-old students, while the Hanover Research (2015) stated that class size reduction can be effective, but mostly in early grades. An explanation for the absence of any 'significant' effect of the percentage of non-native speakers in a school could be the inclusion of the school SES effect: the effect of the % of non-native students and students with a migration background at schools is, as proposed by Gaey et al. (2013), partly captured in the coefficient of the school SES.

5.2.2. System-level characteristics

In general, only one clear correlation was observed in our analyses (Table 4, Models 3C, 4C): the socioeconomic and/or language-based student-inequity is less pronounced in late-tracking (≥ 15) education systems than in early-tracking (<15) education systems. The other system-level characteristics – the incidence of grade repetition, GDP per capita, the level of competitiveness, and the government expenditure on primary education – showed no or only a very weak correlation with the socioeconomic and/or language-based student-inequity.

Concerning the tracking age, we observed that late-tracking education systems show considerably lower levels of socioeconomic and language based student-inequity than early-tracking education systems. These results are in line with the existing literature: the later students are tracked, the higher the student-equity (Lavrijsen & Nicaise, 2015, 2016; Schütz et al., 2008; Brunello & Checchi, 2007; Horn, 2009; Le Donné, 2014; Schlicht et al., 2010; Dupriez et al., 2015). Within the literature, these effects are often explained by the fact that younger students' orientation tends to be more heavily influenced by their role models, such as their parents. In that sense, it can be expected that students' socioeconomic background is more of a burden when tracking decisions are made at younger ages (Lavrijsen, 2015). Where students are tracked at later ages, they have a better idea of what they are capable of and what their interests are.

As regards the frequency of grade repetition, we found a negative correlation with socioeconomic as well as language-based student-inequity. However, all correlations are so small that they actually do not lead to any substantial change in the level of student-equity. Hence, based on our results, we could say that the practice of grade repetition is not an important determinant of student-inequity. This is not in line with previous studies which showed that countries with extensive use of grade repetition face higher levels of student-inequity levels (OECD, 2011b). However, these studies showed that the practice of grade repetition has long-term negative financial, social and academic effects on students. By using PISA data, we are only able to measure the student performance of 15-year-olds. Other long-term effects cannot be measured. Consequently, it is possible that the harmful effects of grade repetition appear in later outcomes such as early school leaving (Jacob and Lefgren, 2009; Yamamoto and Byrnes, 1987; Anderson, Jimerson and Whipple, 2005).

Similarly, we do not find any - or just a very weak - significant correlation between the level of competition between schools and the language based and/or socioeconomic student-inequity, indicating that the level of competitiveness between schools is not related to student-inequity. This is not in line with the large body of the existing literature (Ladd and Walsh, 2002; Ladd, 2003; Soderstrom and Uusitalo, 2005; Field et al, 2007; Ladd, et al., 2011; Musset, 2012; Cullen, 2005; Garcia, 2008; Böhlmark, 2007): it is often suggested that more competitiveness between schools can boost student performance – which we also observed, although it was only a small effect – but also exacerbates student-inequity. A possible explanation for the absence of a correlation could lie in our measure of competitiveness.

Finally, two system-level characteristics did not show any significant correlation with student-equity: GDP per capita and government expenditure on education. This indicates that GDP per capita as well as the government expenditure on primary education have the same impact on learning outcomes of privileged and disadvantaged students. Results concerning the GDP per capita are again illuminating. No direct evidence exists on how the GDP per capita relates to the socio-economic and student-equity. Based on our analyses, we could assume it does correlate with student performance, but not with student-equity. The absence of an association between student-equity and government expenditure on primary education is more surprising. Based on existing studies we had expected to find a negative coefficient (Strietholt, 2019; Schlicht et al., 2010; Salehi-sfahani et al., 2013; Akiba et al., 2007; Bodovski et al., 2017).

5.3. Which system-level characteristics are correlated with *socioeconomic and language-based school-equity* and how?

In order to answer this last research question, we included cross-level interaction terms between our two measures of school-equity –school SES- and the % of non-native speakers– with all system-level characteristics in Models 5C and 6C.

5.3.1. System-level characteristics

In general, we did not find any clear interactions between our system-level characteristics and the level of socioeconomic and language-based school-inequity. We only observe a few small correlations that are significant at the 10%-level, but they do not affect the school-inequity substantially. More specifically, we find a weakly positive association between an education system's level of competitiveness between schools and the socioeconomic school-inequity. Furthermore, we observe a correlation between the level of competitiveness and the language-based school-equity in schools, but this coefficient is almost equal to zero.

In sum, we would conclude that our system-level characteristics do not impact the school-inequity substantially. This is not shocking, given that our system-level characteristics – except the tracking age – were also not related to the student-inequity. This raises the question whether or not we should try to include other system-level characteristics that are more suited in explaining differences in student- and school-inequity.

Conclusion

Given the growing social and economic inequalities in most countries of the world, improving equity in education has become a more urgent matter than ever before (OECD, 2018). During the past few decades, many countries developed several policies to improve the level of performance and equity. Despite these efforts, international studies show that no country in the world can yet claim to have eliminated social and language-based inequities in education (OECD, 2018). There are, however, countries that succeed better than others in improving equity in education, and within countries, some schools are strikingly more equitable than others (OECD 2018; Cumberworth et al., 2018; Gustafsson et al., 2018). These differences between countries and schools prove that there is nothing inevitable about disadvantaged students performing worse than their more privileged counterparts (OECD, 2018).

Within the educational field, however, it is often feared that improving educational equity goes at the expense of average performance. This is part of a wider debate on the so called 'efficiency-equity trade-off'. Several studies suggest that this is a false dilemma (Woesmann et al., 2006): countries can promote efficiency and equity simultaneously. Our study largely supports this conclusion.

Our model was specified in a multilevel framework so as to distinguish between determinants of educational outcomes at student, school and system (or country) level. Moreover, the main effects at system level reflect the average performance ('efficiency') effects, while the cross-level interaction effects reflect the effects on inequalities by socio-economic background and migration status ('equity effects'). Consequently, we were able to analyse how and to what extent student and school performance, as well as student- and school-inequity are related with school- and system-level characteristics.

Based on our analyses, we can highlight four key findings:

1. The (cumulative) school SES is very strongly correlated with student and school performance, indicating that it is a strong determinant of inequities between students and schools: students attending an advantaged school will – regardless of their own socioeconomic background – substantially outperform students attending a disadvantaged school. Conversely, attending a low SES school results in low achievement for all students (even those from high socioeconomic background). Contrary to most expectations, we observe a positive correlation between the (cumulative) school SES and socioeconomic student-inequity. This indicates that advantaged schools are confronted with higher levels of socioeconomic student-inequity than disadvantaged schools. Our findings point at the risk that school segregation reinforces socioeconomic inequalities in performance. Governments should therefore aim at minimising school segregation through direct or indirect measures. Moreover, promoting more socially mixed schools is in the interest of policy-makers as it will improve equity, without significantly affecting the average student performance.

2. Teacher quality – measured as the percentage of fully certified teachers in a school – is positively correlated with student and school performance: the higher the quality of teachers, the better the student and school performance. Moreover, our findings also suggest that more certified teachers in a school, reduces the language-based student-inequities. Hence we can conclude that ensuring a sufficient supply of qualified teachers remains an important responsibility for governments, especially given the fact that it can simultaneously improve the level of performance and reduce educational inequity.
3. Late-tracking education systems show considerably lower levels of socioeconomic and language based student-inequity than early-tracking education systems. At the same time, our results suggest that the average mathematics performance of students in late-tracking education systems is lower than in early-tracking education systems. However, this coefficient needs to be interpreted with great care given that we included non-OECD (i.e. less wealthy) countries and that the coefficients are barely significant at 10%-level. Moreover, the negative effect on performance applies only to maths, not on reading. In earlier studies, Lavrijsen and Nicaise (2015; 2016) corrected for performance levels in grade 4 of primary school and found positive rather than negative effects on average performance, for maths as well as science and reading. Only the top quartile of best performers in maths appeared to have an (insignificant) advantage in early tracking systems. Hence, we can plausibly argue that tracking students at later ages reduces the socio-economic as well as ethnic inequities considerably, while boosting average performance.

One of the motives for this analysis was to frame the discussion about equal opportunity policies – and equity funding in particular – in a broader, systemic context. Indeed, the effectiveness of additional funding targeted at schools with a disadvantaged student population depends on the relative weight of such measures, compared with other system characteristics that may reinforce or hamper their effects.

1. The main message from this analysis is the strong connection between school SES and individual educational opportunities. Any given distribution of student SES can be ‘amplified’ by segregation mechanisms that sort low- and high-SES students into different schools, resulting in more unequal outcomes. What is more, there is a positive interaction between individual SES and school-SES, which makes social elite schools more inegalitarian than average- or low-SES schools. This has three direct implications for equity funding schemes: (a) the phenomenon of social segregation enhances the need for equity funding, as social inequality boosts unequal outcomes. (b) given this ‘amplifier effect’ of school segregation, it is understandable that a given investment in equity funding is less effective in a segregated school landscape than in a context where all schools are socially and language-wise more mixed. (c) If equity funding is used as an instrument to iron out performance gaps between segregated schools, it seems justified to ‘amplify’ the funding per disadvantaged student with a coefficient that varies with average school-SES. For example, suppose that there are two schools that have 200 disadvantaged students each; and suppose that these students represent 80% of the student population in school A and 50% in school B, it would (ceteris paribus) be justified to allocate more equity funding to school A.
2. The argument above also means that, in a way, equity funding plays a more remedial role, aimed at redressing inequalities that are generated by other mechanisms. Preventing school segregation (e.g. by regulating the allocation of students over schools) could reduce the budget cost of equity. Admittedly, regulation appears to meet a lot of resistance in countries

(such as Belgium) where freedom of education and free school choice are seen as constitutional rights. It is important to realise that this freedom comes at a cost. Similar tensions arise in the debate about later tracking (which could also reduce social and ethnic segregation).

3. Recent research (OECD, 2014; OECD, 2018b) has suggested that the 'quality' of teachers (measured by their level of qualification) is more important than their 'quantity' (measured by the teacher/student ratio) in raising the performance of (disadvantaged) students. This would mean that equity funding schemes should foster the recruitment and retention of better trained teachers, rather than just extending the number of teachers, e.g. to reduce class size. The findings in this paper tend to point into the same direction, as the effect of the student-teacher ratio at school level does not seem to substantially influence student outcomes – neither on average, nor its distribution. However, the average student-teacher ratio at school level is a fairly rough measure. A more detailed analysis would be needed to determine the optimal quantity-quality mix, so as to maximise the return on investment.

Our study is not without limitations. First off all, as we have mentioned earlier, PISA data are cross-sectional, with no measure of student skills prior to entering the current school. Consequently, results are not corrected for unequal 'starting positions'. This means that causality cannot be strictly proven. Second, PISA data contain no measure of long-term outcomes. Test scores may be good - though certainly imperfect - predictors of later achievement. We did not always find significant correlations, but it could be that other long-term outcomes will be affected by some school and/or system-level characteristics. Third, the fact that we do not include all school- and system-level factors at the same time in one model (due to some methodological issues) may cause so-called 'omitted variable bias'.

ANNEXES

Annex 1. The Flemish Case

Table 6. School-level characteristics and student performance in the Flemish Community of Belgium

	Model 7C: Maths		Model 8C: Reading	
	B	**SE**	**B**	**SE**
Intercept	466.61***		492.94***	2.80
Level 1				
Age	8.73***	1.58	9.11***	1.61
Migration	-6.78***	1.90	-5.20*	2.26
Language	-14.89***	2.13	-20.55***	2.08
Gender	15.56***	0.86	-18.85***	0.89
SES	14.06***	0.53	13.57***	0.60
Level 2				
% cert teacher	40.06***	7.53	35.90***	7.37
ST/TE-ratio	-1.08**	0.41	-1.47***	0.38
M_SES	54.96***	2.96	64.11***	3.06
M_LANG	-0.67***	0.13	-0.38***	0.14
Random slope				
SES slope	55.04***	11.66	41.55***	13.16
Lang slope	170.66***	79.97	426.19***	13.16
Cross-level interactions				
M_SES*SES	6.43***	0.74	6.28***	0.70
M_LANG*SES	-0.10*	0.04	-0.07	0.05
% CERT TEA*SES	2.49	1.68	2.03	1.85
ST/TE-ratio *SES	0.03	0.11	-0.05	0.10
M_SES*LANG	-3.15†	1.79	-6.59***	2.22
M_LANG*LANG	0.20*	0.09	0.03	0.10
% CERT TEA*LANG	0.20	4.79	0.48	5.89
ST/TE-ratio*LANG	0.69*	0.32	1.36***	0.33

***p<0.001, **p<0.01, *p<0.05, † p<0.1

References

Abdulkadirog˘ lu, A., P. Pathak and C. Walters (2018), “Free to Choose: Can School Choice Reduce Student Achievement?”, American Economic Journal: Applied Economics, Vol. 10/1, pp. 175-206, http://dx.doi.org/10.1257/app.20160634.

Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. Educational Researcher, 36(7), 369–387. https://doi.org/10.3102/0013189x0 7308739 .

Agirdag, O., Van Houtte, M., & Van Avermaet, P. (2012). Why Does the Ethnic and Socio-economic Composition of Schools Influence Math Achievement? The Role of Sense of Futility and Futility Culture. European Sociological Review, 28(3), 366-378. doi:10.1093/esr/jcq070

Ainley, John & Ainley, John, 1945- & Australia. Department of Employment, Education and Training & Australian Council for Educational Research (1995).Socioeconomic status and school education. Australian Govt. Pub. Service, Canberra

Ammermueller, A., & Pischke, J.-S. (2009). Peer effects in European primary schools: Evidence from the progress in international reading literacy study. Journal of Labor Economics, 27(3), 315-348.

An, Brian P. “The Impact of Dual Enrollment on College Degree Attainment: Do Low-SES Students Benefit?” Educational Evaluation and Policy Analysis 35, no. 1 (2013): 57–75

Andersson, C. (2007). Teacher density and student achievement in Swedish compulsory schools.

Bedard, K., & Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. The Quarterly Journal of Economics, 121(4), 1437-1472.

Bellens, K., Arkens, T., Van Damme, J., & Gielen, S. (2013a). Sociale ongelijkheid en ongelijkheid op basis van thuistaal inzake wetenschapsprestaties in het Vlaamse onderwijs. Veranderingen tussen 2003 en 2011 op basis van TIMSS, vierde leerjaar.

Bellens, K., Arkens, T., Van Damme, J., & Gielen, S. (2013b). Sociale ongelijkheid en ongelijkheid op basis van thuistaal inzake wiskundeprestaties in het Vlaamse onderwijs. Veranderingen tussen 2003 en 2011 op basis van TIMSS, vierde leerjaar.

Betts, J., & Loveless, T. (Eds.). (2005). Getting Choice Right: Ensuring Equity and Efficiency in Education Policy. Washington, D.C.: Brookings Institution Press. Retrieved from http://www.jstor.org/stable/10.7864/j.ctt12879sq

Betts, J., & Roemer, J. E. (2005). Equalizing opportunity for racial and socioeconomic groups in the United States through educational finance reform. Department of Economics, UCSD.

Berger, A., Turk-Bicakci, L., Garet, M., Song, M., Knudson, J., Haxton, C., ... & Keating, K. (2013). Early College, Early Success: Early College High School Initiative Impact Study. American institutes for research.

Bjorklund, A., Clark, M. A., Edin, P.-A., Fredricksson, P., & Krueger, A. B. (2005). The market comes to education in Sweden: an evaluation of Sweden's surprising school reforms: Russell Sage Foundation.

Bodovski, K., Byun, S. Y., Chykina, V., & Chung, H. J. (2017). Searching for the golden model of education: Cross-national analysis of math achievement. Compare, 47(5), 722–741. https://doi.org/10.1080/03057925.2016.1274881 .

Boeskens, L. (2016), “Regulating Publicly Funded Private Schools: A Literature Review on Equity and Effectiveness”, OECD Education Working Papers, No. 147, OECD Publishing, Paris, https://dx.doi.org/10.1787/5jln6jcg80r4-en.

Böhlmark, A., & Lindahl, M. (2007). The Impact of School Choice on Pupil Achievement

Bouguen, A., J. Grenet and M. Gurgand (2017), "La taille des classes influence-t-elle la reussite scolaire ?", Les notes de l'IPP, No. 28, Institut des Politiques Publiques, Paris, www.ipp.eu/wp-content/ uploads/2017/09/n28-notesIPP-sept2017.pdf (accessed on 01 December 2017).

Brophy, J. (2006), "Grade Repetition", Education Policy Series No. 6, International Institute for Educational Planning, UNESCO, Paris.

Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. Economic Policy, 22(52), 781–861. doi: http://www.jstor.org/stable/4502215

Burke, M. A., & Sass, T. R. (2013). Classroom peer effects and student achievement. Journal of Labor Economics, 31(1), 51-82

Caena, F. (2011). Literature review Quality in Teachers' continuing professional development. Education and training, 2020, 2-20.

Chiu, M. M. (2015). Family inequality, school inequalities, and mathematics achievement in 65 countries: Microeconomic mechanisms of rent seeking and diminishing marginal returns. Teachers College Record, 117(1), 1–32.

Clotfelter, C., H. Ladd and J. Vigdor (2007), "Teacher credentials and student achievement: Longitudinal analysis with student fixed effects", Economics of Education Review, Vol. 26/6, pp. 673-682, http:// dx.doi.org/10.1016/J.ECONEDUREV.2007.10.002.

Clotfelter, C., H. Ladd and J. Vigdor (2010), "Teacher Credentials and Student Achievement in High School", Journal of Human Resources, Vol. 45/3, pp. 655-681, http://dx.doi.org/10.3368/jhr.45.3.655.

Coleman, J. (1968). The Concept of Equality of Educational Opportunity. Harvard Educational Review, 38(1), 7-22. doi:10.17763/haer.38.1.m3770776577415m2

Crawford, C., Dearden, L., & Meghir, C. (2007). When you are born matters: The impact of date of birth on child cognitive outcomes in England.

Cullen, J. B., Jacob, B. A., & Levitt, S. D. (2005). The impact of school choice on student outcomes: an analysis of the Chicago Public Schools. Journal of Public Economics, 89(5-6), 729-760.

Cumberworth, S., & Cumberworth, E. (2018). Does School Composition Matter More for Lower-SES Students? A Cross-National Examination of School Socioeconomic Composition, Individual Socioeconomic Status, and Standardized Test Scores.

Darling-Hammond, L. (2004), "Inequality and the Right to Learn: Access to Qualified Teachers in California's Public Schools", Teachers College Record, Vol. 106/10, http://internationalteachercert.wiki. educ.msu.edu/file/view/Darling-Hammond+%282004%29.pdf (accessed on 07 December 2017), pp. 1936-196

Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., & Televantou, I. (2018). Effects of school-average achievement on individual self-concept and achievement: Unmasking phantom effects masquerading as true compositional effects. Journal of Educational Psychology, 110(8), 1112.

Ding, W., & Lehrer, S. F. (2010). Estimating treatment effects from contaminated multiperiod education experiments: the dynamic impacts of class size reductions. The Review of Economics and Statistics, 92(1), 31-42.

Dockx, J. (2019), *On track? A study on the effects of tracks in secondary education*, KU Leuven: PhD dissertation

Driessen, G. (2002). School composition and achievement in primary education: A large-scale multilevel approach. Studies in Educational Evaluation, 28(4), 347-368.

Dumay, X., & Dupriez, V. (2008). Does the school composition effect matter? Evidence from Belgian data. British Journal of Educational Studies, 56(4), 440-477.

Ecalle, J., Magnan, A., & Gibert, F. (2006). Class size effects on literacy skills and literacy interest in first grade: A large-scale investigation. Journal of School Psychology, 44(3), 191–209.

Edmunds, J. A., Unlu, F., Glennie, E., Bernstein, L., Fesler, L., Furey, J., & Arshavsky, N. (2017). Smoothing the transition to postsecondary education: The impact of the early college model. Journal of Research on Educational Effectiveness, 10(2), 297-325.

European Commission/EACEA/Eurydice. (2014). Financing Schools in Europe: Mechanisms, Methods and Criteria in Public Funding. Luxembourg: Publications Office of the European Union.

European Commission/EACEA/Eurydice. (2015). The teaching profession in Europe: practices, perceptions, and policies. Luxembourg: Publications Office of the European Union.

Epple, D., R. Romano and M. Urquiola (2017), "School Vouchers: A Survey of the Economics Literature", Journal of Economic Literature, Vol. 55/2, pp. 441-492, http://dx.doi.org/10.1257/jel.20150679

Franck, Emilie; Nicaise, Ides; 2019. Ongelijkheden in het Vlaamse onderwijssysteem: verbetering in zicht? Een vergelijking tussen PISA 2003 en 2015 Publisher: HIVA / SONO (Steunpunt Onderwijsonderzoek); Leuven.

Fekjær, S. N., & Birkelund, G. E. (2007). Does the ethnic composition of upper secondary schools influence educational achievement and attainment? A multilevel analysis of the Norwegian case. European Sociological Review, 23(3), 309-323

Garcia, D. R. (2008). The impact of school choice on racial segregation in charter schools. Educational Policy, 22(6), 805-829.

Geay, C., McNally, S., & Telhaj, S. (2013). Non-native speakers of English in the classroom: what are the effects on pupil performance?. The Economic Journal, 123(570), F281-F307.

Gustafsson, J. E., Nilsen, T., & Hansen, K. Y. (2018). School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011. *Studies in Educational Evaluation*, 57, 16-30.

Gylfason, T., & Zoega, G. (2003). Education, social equality and economic growth: a view of the landscape. CESifo Economic Studies, 49(4), 557-579.

Hanushek, E. A. (2003). The failure of input-based schooling policies. The economic journal, 113(485), F64-F98.

Hanushek, E. A. (2006). School resources. Handbook of the Economics of Education, 2, 865-908.

Heckman, J. J. (2008). Schools, Skills, and Synapses. Econ Inq, 46(3), 289.

Heyerick, L. (1978), *De schoolse achterstand van arbeiderskinderen. Een onderzoek naar het aandeel van de leerkracht-leerlingrelatie in het totstandkomen van die problematiek*, UGent: PhD dissertation

Horn, D. (2009). Age of selection counts: A cross-country analysis of educational institutions. Educational Research and Evaluation, 15(4), 343–366. https://doi.org/10.1080/13803610903087011

Hout, M., & DiPrete, T. A. (2006). What we have learned: RC28's contributions to knowledge about social stratification. Research in social stratification and mobility, 24(1), 1-20.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation.

Hoxby, C. (2002), School Choice and School Productivity (or Could School Choice be a Tide that Lifts All Boats?), National Bureau of Economic Research, Cambridge, MA, http://dx.doi.org/10.3386/w8873.

Jacobs, D., Rea, A., & Hanquinet, L. (2007). Prestaties van de leerlingen van buitenlandse herkomst in België volgens de PISA-studie: vergelijking tussen de Franse Gemeenschap en de Vlaamse Gemeenschap. Brussel: Koning Bouwdewijnstiching.

Jahanshahi, B. ., & Naghavi, A. . (2017). Education reform and education gaps. Applied Economics Letters, 24(19), 1–4.

Jensen, P., & Rasmussen, A. W. (2011). The effect of immigrant concentration in schools on native and immigrant children's reading and math skills. Economics of education Review, 30(6), 1503-1515.

Jimerson, S., G. Anderson and A. Whipple (2002), "Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school", Psychology in the Schools, Vol. 39.

Jimerson, S., S. Pletcher and K. Graydon (2006), "Beyond grade retention and social promotion: Promoting the social and academic competence of students", Psychology in the schools, Vol. 43, No.1, pp. 85-97.

Krueger, A. B., & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. The Economic Journal, 111(468), 1-28.

Ladd, H. F., & Walsh, R. P. (2002). Implementing value-added measures of school effectiveness: getting the incentives right. Economics of Education review, 21(1), 1-17.

Ladd (2003) Comment on Caroline M. Hoxby: School choice and school competition: Evidence from the United States

Ladd, H. F., Fiske, E. B., & Ruijs, N. (2011). Does parental choice foster segregated schools? Insights from the Netherlands. In M. Berends, M. Cannata, & E. B. Goldring (Eds.), School choice and school improvement (pp. 233 - 253): Harvard Education Press.

Lavrijsen, J., & Nicaise, I. (2015). New empirical evidence on the effect of educational tracking on social inequalities in reading achievement. European Educational Research Journal, 14(3–4), 206-221. https://doi.org/10.1177/1474904115589039

Lavrijsen, J., & Nicaise, I. (2016). Educational tracking, inequality and performance: New evidence from a differences-in-differences technique, *Research in Comparative & International Education*, Vol. 11(3) 334–349

Lavy, V., Silva, O., & Weinhardt, F. (2012). The good, the bad, and the average: Evidence on ability peer effects in schools. Journal of Labor Economics, 30(2), 367-414.

Le Donné, N. (2014). European variations in socioeconomic inequalities in students' cognitive achievement: The role of educational policies. European Sociological Review, 30(3), 329–343. https://doi. org/10.1093/esr/jcu040 .

Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2007). The effect of extra funding for disadvantaged pupils on achievement. The Review of Economics and Statistics, 89(4), 721-736.

Leuven, E., Oosterbeek, H., & Rønning, M. (2008). Quasi-experimental estimates of the effect of class size on achievement in Norway. The Scandinavian Journal of Economics, 110(4), 663-693.

Li, X. (2010). How unequal are their vertical equity measures? A comparison of three Canadian provinces. Alberta Journal of Educational Research, 56(2), 218

Li, W., & Konstantopoulos, S. (2016). Class Size Effects on Fourth-Grade Mathematics Achievement: Evidence From TIMSS 2011. Journal of Research on Educational Effectiveness, 9(4), 503-530.

Luschei, T. F. & Chudgar, A. (2011). Teachers, student achievement and national income: A cross-national examination of relationships and interactions. Prospects, 41, 507– 533

Matthew Gianni, Celeste Alexander, and Pedro Reyes. "Exploring Variation in the Impact of Dual-Credit Coursework on Postsecondary Outcomes: A Quasi-Experimental Analysis of Texas Students." The High School Journal 97, no. 4 (2014): 200–218

McConney, A., & Perry, L. B. (2010a). Science and mathematics achievement in Australia: The role of school socioeconomic composition in educational equity and effectiveness. International Journal of Science and Mathematics Education, 8(3), 429-452.

McKinsey, Barber, M., & Mourshed, MonaMcKinsey, Barber, M., & Mourshed, M. (2007). How the world's best-performing school systems come out on top. Analysis, 18(3), 56.

Monk, D. (1994), "Subject area preparation of secondary mathematics and science teachers and student achievement", Economics of Education Review, Vol. 13/2, pp. 125-145, http://dx.doi.org/10. 1016/0272-7757(94)90003-5.

Montes, O., & Rubalcaba, L. (2014). School choice, equity and efficiency: International evidence from PISA-2012. Asociación de Economía de la Educación, 1-35.

Musset, P. (2012), “School Choice and Equity: Current Policies in OECD Countries and a Literature Review”, OECD Education Working Papers, No. 66, OECD Publishing, Paris, https://dx.doi.org/10.1787/5k9fq23507vc-en.

Nusche, D. (2009). What Works in Migrant Education? A Review of Evidence and Policy Options. OECD Education Working Papers, No. 22. OECD Publishing.

Nusche, D., Miron, G., Santiago, P., & Teese, R. (2015). OECD Reviews of School Resources: Flemish Community of Belgium 2015.

OECD. (2005). Teachers matter: Attracting, developing and retaining effective teachers: OECD publishing.

OECD (2009). PISA Data Analysis Manual, SPSS Second Edition, OECD Publishing, Paris.

OECD. (2011a). Divided We Stand: Why Inequality Keeps Rising. Paris: OECD Publishing.

OECD (2011b), Lessons from PISA for the United States, Strong Performers and Successful Reformers in Education, OECD, Paris.

OECD (2012c), "Does Money Buy Strong Performance in PISA?", PISA in Focus, No. 13, OECD Publishing, Paris, https://doi.org/10.1787/5k9fhmfzc4xx-en.

OECD. (2012d). Equity and Quality in Education. Paris: OECD Publishing.

OECD. (2013a). PISA 2012 Results: Excellence through Equity (Volume II). Paris: OECD Publishing.

OECD (2013b), PISA 2012 Results: What Makes Schools Successful (Volume IV): Resources, Policies and Practices, PISA, OECD Publishing, Paris.

OECD (2015a), Immigrant Students at School: Easing the Journey towards Integration, OECD Reviews of Migrant Education, OECD Publishing, Paris, https://doi.org/10.1787/9789264249509-en.

OECD. (2015C). The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence. Paris: OECD Publishing.

OECD. (2015d). In It Together: Why Less Inequality Benefits All. Paris: OECD Publishing.

OECD. (2016a). PISA 2015 Results (Volume I): Excellence and Equity in Education. Paris: OECD Publishing.

OECD (2016B), PISA 2015 Results (Volume II): Policies and Practices for Successful Schools, PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264267510-en.

OECD. (2016). PISA 2015 Technical Report. Paris: OECD Publishing.

OECD. (2017b). Empowering and Enabling Teachers to Improve Equity and Outcomes for All. Paris: OECD Publishing.

OECD, (2017C). The Funding of School Education. OECD Publishing. Paris.

OECD (2017E), *School choice and school vouchers: An OECD perspective*, www.oecd.org/education/School-choice-and-school-vouchers-an-OECD-perspective.pdf

OECD (2018). Equity in Education: Breaking Down Barriers to Social Mobility. Paris: OECD Publishing.

OECD (2018b), Effective Teacher Policies: Insights from PISA, PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264301603-en.

OECD. (2019). Balancing School Choice and Equity: An International Perspective Based on Pisa. Paris: OECD Publishing.

Palardy, G. J. (2008). Differential school effects among low, middle, and high social class composition schools: A multiple group, multilevel latent growth curve analysis. School Effectiveness and School Improvement, 19(1), 21-49.

Pelleriaux, K. *Demotie en burgerschap. De culturele constructie van ongelijkheid in de kennismaatschappij,* VUBPRESS, Brussel, 2001

Perry, L. B. (2007). Conceptualizing education policy in democratic societies. Educational Policy, 23(3), 423-450.

Perry, L. (2009). Characteristics of equitable systems of education: A cross-national analysis. European Education, 41(1), 79-100.

Perry, L. B., & McConney, A. (2010b). Does the SES of the school matter? An examination of socioeconomic status and student achievement using PISA 2003. Teachers College Record, 112(4), 1137-1162.

Perry, L. B., & McConney, A. (2013). School socioeconomic status and student outcomes in reading and mathematics: A comparison of Australia and Canada. Australian Journal of Education, 57(2), 124-140.

Pfeffer, F. T. (2008). Persistent inequality in educational attainment and its institutional context. European sociological review, 24(5), 543-565.

Poesen-Vandeputte, M., Nicaise, I. (2015), Rich schools, poor schools. Hidden resource inequalities between primary schools, Educational Research, 57(1), p.91-109

Roemer, J. E. (1998). Equality of opportunity. Cambrigde Harvard University Press.

Rosenthal, R., Jacobson, L. (1968), Pygmalion in the classroom, *The Urban Review*, 3, pp. 16-20

Ronfeldt, M. and M. Reininger (2012), "More or better student teaching?", Teaching and Teacher Education, Vol. 28/8, pp. 1091-1106, http://dx.doi.org/10.1016/J.TATE.2012.06.003.

Rumberger, R. W., & Palardy, G. J. (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. American educational research journal, 42(1), 3-42.

Salehi-Isfahani, D., Hassine, N., & Assaad, R. (2013). Equality of opportunity in educational achievement in the Middle East and North Africa. The Journal of Economic Inequality, 12(4), 489–515. https://doi.org/10.1007/s10888-013-9263-6 .

Schütz, G., Ursprung, H. W., & Woessmann, L. (2008). Education policy and equality of opportunity. KyKlos, 61(2), 279–308. https://doi.org/10.1111/j.1467-6435.2008.00402.x .

Schlicht, R., Stadelmann-Steffen, I., & Freitag, M. (2010). Educational inequality in the EU the effectiveness of the national education policy. European Union Politics, 11(1), 29–60. https://doi.org/10.1 177/1465116509346387 .

Shen, T., & Konstantopoulos, S. (2017). Class size effects on reading achievement in europe: evidence from pirls. Studies in Educational Evaluation, 53, 98-114.

Söderström, M., & Uusitalo, R. (2010). School Choice and Segregation: Evidence from an Admission Reform. The Scandinavian Journal of Economics, 112(1), 55-76. Retrieved from http://www.jstor.org/stable/40587796

Solli, I. F. (2017). Left behind by birth month. Education Economics, 25(4), 323-346.

Speybroeck, S. (2013), *Teachers' expectations and the achievement gap: the role of students' ethnicity and socio-economic status*, KU Leuven: PhD dissertation,

Spruyt, B. (2014), De onderwaardering van technisch en beroepsonderwijs: oorzaken en gevolgen, in: Nicaise, I., Spruyt, B., Van Houtte, M., Kavadias, D. (eds.,2014), *Het onderwijsdebat. Waarom de hervorming van het secundair broodnodig is*, Berchem: EPO, p.71-88

Strietholt, R., Gustafsson, J.-E., Hogrebe, N., Rolfe, V., Rosén, M., Steinmann, I., & Yang Hansen, K. (accepted). The Impact of Education Policies on Socioeconomic Inequality in Student Achievement: A Review of Comparative Studies. In L. Volante, S. Schnepf, J. Jerrim, & D. Klinger (eds.): *Socioeconomic Inequality and Student Outcomes: National Trends, Policies, and Practices* (pp. xx–xx). New York, NY: Springer.

Struhl, B., & Vargas, J. (2012). Taking College Courses in High School: A Strategy Guide for College Readiness--The College Outcomes of Dual Enrollment in Texas. Jobs for the Future.

Tarnowski, K. J., Anderson, D. F., Drabman, R. S., & Kelly, P. A. (1990). Disproportionate referrals for child academic/behavior problems: Replication and extension. Journal of Consulting and Clinical Psychology, 58(2), 240.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research? *Review of Educational Research*, 83(3), 357–385.

Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research*, 37(5), 483–504.

Tonello, M. (2016). Peer effects of non-native students on natives' educational outcomes: mechanisms and evidence. Empirical Economics, 51(1), 383-414.

Urquiola, M. (2016), "Competition Among Schools: Traditional Public and Private Schools",Handbook of the Economics of Education, Vol. 5, pp. 209-237, http://dx.doi.org/10.1016/B978-0-444-63459-7.00004-X.

U.S. Department of Education (2017). Institute of Education Sciences. "Transition to College Intervention Report: Dual Enrollment Programs." What Works Clearninghouse, February 2017. https://ies.ed.gov/ncee/wwc/Intervention/1043.

Vakgroep Onderwijskunde. (2015). *WETENSCHAPPELIJKE GELETTERDHEID BIJ 15-JARIGEN: VLAAMS RAPPORT PISA 2015*.

Van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. Educational Research Review, 5(2), 134-150.

Vandenbroeck, M., Vanlaar, G., Bellens, K., Van Damme, J., & De Fraine, B. (2016). Het Vlaams lager onderwijs in TIMSS 2015: Wiskunde en wetenschappen in internationaal perspectief en in vergelijking met vorige deelnames. Leuven: KU Leuven, Centrum voor Onderwijseffectiviteit en -evaluatie.

Van der Slik, F. W., Driessen, G. W., & De Bot, K. L. (2006). Ethnic and socioeconomic class composition and language proficiency: a longitudinal multilevel examination in Dutch elementary schools. European Sociological Review, 22(3), 293-308.

Van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. Educational Research Review, 5(2), 134– 150.

Verachtert, P., De Fraine, B., & Ghesquière, P. (2008). De relatie tussen geboortekwartaal en schools succes in de eerste jaren van het lager onderwijs. Hasselt: Vlaams Forum voor Onderwijsonderzoek.

Van Houtte, M. & Stevens, P.A.J. 'The culture of futility and its impact on study culture in technical/vocational schools in Belgium', in: *Oxford Review of Education*, 2010, 36(1), 23-43.

Wallingford, E. L., & Prout, H. T. (2000). The relationship of season of birth and special education referral. Psychology in the Schools, 37(4), 379-387.

Wells, R. (2010). Children of Immigrants and Educational Expectations: The Roles of School Composition. Teachers College Record, 112(6), 1679-1704.

Willms, J. D. (1986). Social class segregation and its relationship to pupils' examination results in Scotland. American sociological review, 224-241.

Wilson, G. (2000). The effects of season of birth, sex and cognitive abilities on the assessment of special educational needs. Educational Psychology, 20(2), 153-166.

Woessmann, L. (2009). International evidence on school tracking: A review. CESifo DICE Report, 7(1), 26-34.

Woessmann, L., & Schütz, G. (2006). Efficiency and equity in European education and training systems. Analytical Report for the European Commission prepared by the European Expert Network on Economics of Education,(Bruselas: Comisión Europea, 2006) http://www. eenee. de/portal/page/portal/EENEEContent/_IMPORT_TELECENTRUM/DOCS/SWP% 20EENEE-Part, 20060426.

Wu, M. (2005). The role of plausible values in large-scale surveys. Studies in Educational Evaluation, 31(2-3), 114-128.