

Naar een beter begrip van de intra-klasse correlatie: Een reactie op de discussiebijdrage “Schoolverschillen en schooleffecten in het voorgezet onderwijs”

A.C. Timmermans

Samenvatting

In de discussiebijdrage “Schoolverschillen en schooleffecten in het voorgezet onderwijs” (Lek, Feskens, & Scheerens, 2020) wordt gewezen op enkele voorbeelden waarin onderzoekers of beleidsmakers blijken te hebben gegeven nog onvoldoende grip te hebben op gepresenteerde intra-klasse correlaties. In deze reactie beschrijf ik de noodzaak om een veel explicieter onderscheid te maken tussen de intra-klasse correlatie als een generieke maat voor de mate van afhankelijkheid van hiërarchisch gestructureerde data en het onderliggende statistische meerniveau-model en benoem ik een aantal misinterpretaties van de intra-klasse correlatie uit de discussiebijdrage. Ik ga in op drie thema’s, namelijk: 1) De intra-klasse correlatie is een generieke parameter die gebruikt kan worden voor het beschrijven van variantiecomponenten uit verschillende statistische meerniveau-modellen voor hiërarchisch gestructureerde data. 2) Een intra-klasse correlatie als verhouding tussen variantiecomponenten kan niet rekening houden met eventuele correctiefactoren zoals initiële prestaties, andere kenmerken van leerlingen of schoolsystemen; dat kan alleen in het onderliggende meerniveau-model. 3) De intra-klasse correlatie is geen geschikte maat om de grootte van verschillen in prestaties tussen scholen tussen verschillende landen te vergelijken. Ten slotte roep ik auteurs van onderzoeksverslagen op om veel explicieter te zijn dan tot nu toe gebruikelijk in de specificatie van het onderliggende meerniveau-model. Het is belangrijk om enerzijds duidelijk te maken naar welk hiërarchisch niveau de intra-klasse correlatie verwijst en anderzijds of correctiefactoren in het onderliggende model zijn toegepast en welke dat zijn.

1. Inleiding

In de discussiebijdrage “Schoolverschillen en schooleffecten in het voorgezet onderwijs” (Lek, Feskens, & Scheerens, 2020) wordt gewezen op enkele voorbeelden waarin onderzoekers of beleidsmakers blijken te hebben gegeven nog onvoldoende grip te hebben op gepresenteerde intra-klasse correlaties in de PISA-onderzoeksrapportages (e.g., Inspectie van het Onderwijs, 2017; McKinsey, 2020). In de betreffende discussiebijdrage worden drie beperkingen beschreven die specifiek betrekking hebben op het interpreteren van de intra-klasse correlatie (zoals gerapporteerd in PISA) als indicator voor de mate waarin scholen in het voortgezet onderwijs verschillen in onderwijskwaliteit. Er wordt gesteld dat de intra-klasse correlatie “1) ons niet vertelt waar verschillen tussen scholen vandaan komen, 2) niet zonder meer rekening houdt met initiële niveauverschillen, en 3) in internationale vergelijkingen niet corrigeert voor schoolstelselverschillen” (Lek et al., 2020, p. 326).

De door de auteurs genoemde beperkingen hebben, in tegenstelling tot wat in de discussiebijdrage wordt gesteld, niet te maken met de intra-klasse correlatie op zichzelf, maar deze vloeien voort uit het onderliggende statistische model. Dit blijkt ook uit de door de auteurs gepresenteerde alternatieven die alle betrekking hebben op de specificatie van onderliggende statistische modellen en keuzes die gemaakt kunnen worden omtrent relevante correctiefactoren. Het gaat dan om vragen als: Is er voldoende rekening gehouden met aanvangsprestaties van leerlingen? Is er gecorrigeerd voor eventueel vroege of late selectie in het onderwijssysteem?

De auteurs stellen naar mijn mening terecht, en overeenkomstig met een lange traditie van onderwijs-effectiviteitsonderzoek (e.g., Goldstein, 2001; Hill & Rowe, 1996;

Raudenbush & Bryk, 1986; Timmermans, 2012), dat in de PISA-rapportage onvoldoende rekening is gehouden met dergelijke factoren om te kunnen spreken van verschillen in gerealiseerde onderwijskwaliteit tussen scholen. Omdat we weten dat de verschillen tussen scholen afnemen wanneer adequaat rekening gehouden wordt met verschillen in instroom (Scheerens & Bosker, 1997) en tevens dat de voorspellende waarde van een aantal relevante correctiefactoren tussen landen verschilt (e.g., OECD, 2016, 2019), moeten dergelijke gegevens uit de PISA-rapportage met grote voorzichtigheid geïnterpreteerd worden. Er zijn reeds verschillende beschouwingen geschreven over welke kenmerken van leerlingen en/of scholen bij voorkeur meegenomen zouden moeten worden in de bepaling van netto school effecten (e.g., Darmawan & Keeves, 2006; De Fraine et al., 2002; Timmermans & Thomas, 2015) en welke consequenties deze keuze heeft voor de geschatte toegevoegde waarde van individuele scholen (e.g., Lauder et al., 2010; Peng & Klieme, 2014; Tekwe et al., 2004; Thomas & Mortimore, 1996; Timmermans et al., 2011).

In deze reactie op de discussiebijdrage wil ik de genoemde beperkingen van het onderliggende statistische model en de discussie rondom de keuze voor correctiefactoren bij het bepalen van de toegevoegde waarde van scholen derhalve niet weerleggen. Wel denk ik dat het noodzakelijk is om een veel explicieter onderscheid te maken tussen de intra-klasse correlatie als een generieke maat voor de mate van afhankelijkheid van hiërarchisch gestructureerde data en het onderliggende statistische meerniveau-model. Deze reactie gaat na een korte introductie over intra-klasse correlaties in op de volgende drie aspecten:

1. De intra-klasse correlatie is een generieke parameter die gebruikt kan worden voor het beschrijven van variantiecomponenten uit verschillende statistische meerniveau-modellen voor hiërarchisch gestructureerde data;
2. Een intra-klasse correlatie als verhouding tussen variantiecomponenten kan niet rekening houden met eventuele

correctiefactoren zoals initiële prestaties, andere kenmerken van leerlingen of schoolsystemen; dat kan alleen in het onderliggende meerniveau-model;

3. De intra-klasse correlatie is geen geschikte maat om de grootte van verschillen in prestaties tussen scholen tussen verschillende landen te vergelijken.

2. De intra-klasse correlatie: Een korte introductie

De intra-klasse correlatie is een generieke parameter die, onder andere in onderzoek naar onderwijs-effectiviteitsonderzoek, gebruikt wordt om de mate van afhankelijkheid weer te geven in data met een hiërarchische structuur (e.g., Goldstein, Browne, & Rasbash, 2002; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Het gaat hier om onderzoeksgegevens die zijn verzameld door middel van een hiërarchische of cluster-steekproef. Hierbij wordt bijvoorbeeld eerst een steekproef getrokken van scholen (niveau 2), waarbinnen dan een deel van de leerlingen of alle leerlingen (niveau 1) worden geselecteerd. De afhankelijkheid in de data betekent dat de leerlingen van dezelfde school in prestaties meer op elkaar lijken dan leerlingen uit een volledig willekeurige steekproef, omdat ze voor een deel dezelfde ervaringen, dezelfde leerkrachten en methodes, dezelfde gelegenheid tot leren, etc. hebben gehad.

Bij data met een hiërarchische structuur is het van belang om zogenaamde meerniveau-modellen toe te passen (e.g., Hill & Rowe, 1996; Hox, 2010; Snijders & Bosker, 2012) om te voorkomen dat foutieve inferenties gemaakt worden over de significantie van voorspellende variabelen, bijvoorbeeld aanvangsprestaties of andere kenmerken van leerlingen. In de eerder geschetste situatie met twee niveaus (leerlingen subscript i genest binnen scholen subscript j) wordt in meerniveau-analyse de variantie in de afhankelijke variabele (al dan niet na correctie voor eventuele covariaten zoals aanvangsprestaties) opgedeeld in een deel tussen scholen (σ_{u0}^2) en een deel tussen

leerlingen binnen scholen (σ_{e0}^2).¹ In deze simpelste vorm van meerniveau-analyse wordt de intra-klasse correlatie bepaald als de proportie variantie op schoolniveau ten opzichte van de totale variantie (zie Formule 1).

$$ICC = \frac{\sigma_{u0}^2}{\sigma_{e0}^2 + \sigma_{u0}^2} \quad (1)$$

De intra-klasse correlatie kan geïnterpreteerd worden als de correlatie tussen twee willekeurig gekozen eenheden uit dezelfde eenheid van het hogere niveau. In het eerdere voorbeeld betreft het de correlatie tussen twee willekeurig gekozen leerlingen uit dezelfde school. De intra-klasse correlatie is begrensd tussen 0 en 1, waarbij in zijn algemeenheid geldt: hoe groter de intra-klasse correlatie is, hoe groter de afhankelijkheid is in de data. Een waarde van 0 betekent dat alle variatie in de afhankelijke variabele strikt tussen leerlingen zit en scholen niet van elkaar verschillen. De waarde 1 betekent dat alle variatie in de afhankelijke variabele strikt tussen scholen zit en leerlingen binnen scholen niet van elkaar verschillen. Heeft de ene school veel leerlingen met hoge prestaties en de andere school veel leerlingen met lage prestaties, dan neemt de variantie tussen scholen (σ_{u0}^2) relatief toe en gaat de intra-klasse correlatie omhoog.²

Data met een hiërarchische structuur zijn er in vele verschillende variaties, met twee of meer geneste niveaus (e.g., Leckie, 2013; Timmermans & van der Werf, 2017). Een paar veelvoorkomende voorbeelden zijn: herhaalde metingen (niveau 1) genest binnen leerlingen (niveau 2) genest binnen klassen (niveau 3), of leerlingen (niveau 1) binnen scholen (niveau 2) binnen regio's of lokale bestuurlijke eenheden (niveau 3). Ook kan er sprake zijn van data waarbij er geen strikte nesting is, zoals data met een kruis-klassificatie (e.g., Goldstein & Sammons, 1997; Hill & Goldstein, 1998; Leckie & Goldstein, 2009; Rasbash & Browne, 2001; Timmermans, Snijders, & Bosker, 2013). Een voorbeeld van data met een kruis-klassificatie betreft leerlingen (niveau 1) die genest zijn in scholen (niveau 2) en buurten (niveau 2). Scholen en buurten hoeven niet per se strikt genest te zijn,

aangezien niet alle leerlingen uit een specifieke buurt naar dezelfde school gaan en vice versa.

Ook in de voorgaande, meer complexe, situaties kan de variantie in de afhankelijke variabele worden opgedeeld in kleinere delen die aan specifieke hiërarchische niveaus gerelateerd zijn. Daarom kunnen intra-klasse correlaties ook in deze meer complexe situaties toegepast worden, om de mate van afhankelijkheid tussen de verschillende niveaus weer te geven. Een intra-klasse correlatie kan derhalve refereren naar klassen, buurten, regio's, leerlingen of andere hiërarchische niveaus. Wat de precieze interpretatie van de intra-klasse correlatie is, hangt derhalve zeer af van de veronderstelde datastructuur in het meerniveau-regressie-model.

3. Interpretatie en misinterpretatie van intra-klasse correlaties

3.1 Intra-klasse correlatie is een generieke maat voor mate van afhankelijkheid

“In schooleffectiviteitsonderzoek worden de verschillen tussen scholen als aangeduid door de intra-klasse correlatie ook wel “bruto” schoolverschillen genoemd” (Lek et al., 2020, p. 324). In de discussiebijdrage wordt hiermee de intra-klasse correlatie specifiek gekoppeld aan bruto schoolverschillen die ook gepresenteerd zijn in de PISA-rapportages. De intra-klasse correlatie is echter een generieke parameter die gebruikt wordt in vele verschillende meerniveau-modellen (e.g., Goldstein et al., 2002; Snijders & Bosker, 2012). De interpretatie van de intra-klasse correlatie is daardoor direct afhankelijk van de veronderstelde hiërarchische datastructuur en het is daarom problematisch om de intra-klasse correlatie zonder meer gelijk te stellen aan schoolverschillen. De intra-klasse correlatie kan immers ook verwijzen naar het aandeel van de variantie tussen bijvoorbeeld klassen, buurten, regio's en leerlingen.

3.2 Intra-klasse correlatie kan niet rekening houden met initiële verschillen

“Ondanks deze “state of the art” wordt in de praktijk nog vaak gebruik gemaakt van de

intra-klasse correlatiecoëfficiënt, die louter een soort “all-in” bruto schoolverschillen reflecteert” (Lek et al., 2020, p. 330). De discussiebijdrage lijkt hiermee te suggereren dat de intra-klasse correlatie gebonden is aan een leeg meerniveau-model waarin geen enkele correctie heeft plaatsgevonden, in welk geval de variantie tussen scholen een indicatie geeft van de bruto of ongecorrigeerde schoolverschillen. De intra-klasse correlatie, als verhouding tussen variantiecomponenten op verschillende hiërarchische niveaus, kan eenvoudigweg niet corrigeren voor initiële prestatieverschillen en overige kenmerken van leerlingen die aan de leerontwikkeling gerelateerd zijn (e.g., opleidingsniveau ouders). Echter, ook op basis van meerniveau-modellen waarbij wél voorspellende variabelen worden meegenomen, kan de intra-klasse correlatie bepaald worden. In dat geval weerspiegelt de intra-klasse correlatie de verhouding tussen de componenten van de residuele variantie ($y_{ij}|x_{ij}$) (Goldstein et al., 2002). De intra-klasse correlatie krijgt dan in zekere zin de interpretatie van een partiële correlatie. Indien op adequate manier rekening is gehouden met aanvangsverschillen en eventuele andere kenmerken van leerlingen of de schoolpopulatie, geeft de residuele variantie tussen scholen een indicatie van de netto schooleffecten of toegevoegde waarde (Bosker et al., 2001; Timmermans, 2012). Dat betekent ook dat de intra-klasse correlatie op schoolniveau gebaseerd op de modelsuggesties uit de discussiebijdrage (model netto-1 en model netto-2) een weerspiegeling geeft van netto schooleffecten.

3.3 Geen geschikte maat voor vergelijken van schoolverschillen in internationaal vergelijkend onderzoek

Door landen tegen elkaar af te zetten in termen van intra-klasse correlaties, zoals in Figuur 1 uit de discussiebijdrage is gedaan, wordt gesuggereerd dat de intra-klasse correlatie een valide maat is om de mate van bruto schoolverschillen tussen landen te vergelijken. Een dergelijke figuur impliceert bijvoorbeeld dat de bruto schoolverschillen van IJsland en Finland in 2015 of van

Griekenland en Zwitserland in 2018 nagenoeg gelijk waren. Waar we zeker van mogen zijn is dat, in een twee-niveau-situatie, voor landen met een gelijke intra-klasse correlatie de mate van afhankelijkheid in de data, oftewel de verhouding tussen de variantie op leerling- en schoolniveau, vergelijkbaar is. De vraag die dan rijst is de volgende: zijn de bruto schoolverschillen ook echt vergelijkbaar wanneer de totale variantie in prestaties van leerlingen sterk tussen de landen verschilt? Immers, wanneer we twee landen met dezelfde intra-klasse correlatie vergelijken, zijn de bruto schoolverschillen dan niet groter in het land met de grootste totale variantie in prestaties?

De intra-klasse correlatie is een gestandaardiseerde maat (range 0 - 1), waardoor ze veel gebruikt wordt om de resultaten van verschillende, onafhankelijk uitgevoerde schooleffectiviteitsonderzoeken te vergelijken (e.g., Scheerens & Bosker, 1997). De standaardisatie vindt bij de intra-klasse correlatie plaats op basis van de totale variatie (al dan niet na correctie voor initiële prestaties en andere covariaten). Dit is wenselijk voor het vergelijken van losse onderwijs-effectiviteitsonderzoeken, omdat er veelal sprake is van zeer verschillende manieren om de prestaties van leerlingen in kaart te brengen (e.g., examencijfers, gestandaardiseerde testen, wel/niet diploma behalen), waardoor de variantie tussen studies sterk kan verschillen. Verschillen in de totale variantie tussen deze studies kunnen dan niet zinvol geïnterpreteerd worden. Het gebruiken van de intra-klasse correlatie is dan de enige manier om de studies toch enigszins vergelijkbaar te maken.

Indien er sprake is van gestandaardiseerde metingen van de prestaties van leerlingen tussen de te vergelijken studies of tussen te vergelijken landen, dan is de ruwe mate van tussen-schoolse variantie (σ_{u0}^2) mogelijk-erwijs een betere indicatie van de bruto- of netto schoolverschillen, afhankelijk van de in het model opgenomen covariaten. Dit is bij verschillende internationaal vergelijkende studies aan de orde, denk hierbij aan bijvoorbeeld Trends in Mathematics and Science Study (TIMSS) en Progress in

International Reading Literacy Study (PIRLS). In deze studies zijn prestaties gemeten op een gestandaardiseerde schaal met een gemiddelde van 500 en een standaarddeviatie van 100 welke over landen vergelijkbaar is (e.g., Mullis et al., 2017; Mullis et al., 2020).

Het presenteren van de (residuele) tussen-schoolse variantie heeft, in het geval er sprake is van gestandaardiseerde metingen, een ander voordeel boven de intra-klasse correlatie. De variantie componenten die nodig zijn voor het berekenen van de intra-klasse correlatie, de tussen-schoolse variantie (σ_{u0}^2) en de variantie tussen leerlingen binnen scholen (σ_{e0}^2), worden in de meerniveau modellen voor ieder land met een bepaalde mate van onzekerheid geschat (standaard error). De mate van onzekerheid omtrent de schatting van de (residuele) tussen-schoolse variantie is afhankelijk van een aantal zaken, waaronder het aantal scholen in de steekproef van een land. Hoe meer scholen in de steekproef, hoe nauwkeuriger de (residuele) tussenschoolse variantie geschat kan worden. Wanneer de (residuele) tussen-schoolse variantie wordt gebruikt om bruto of netto school-verschillen weer te geven kan de bijbehorende standaard error relatief eenvoudig worden gebruikt om betrouwbaarheids-intervallen te construeren en derhalve de mate van onzekerheid rond de gevonden schoolverschillen op een adequate manier weer te geven. Hiermee voorkomt men eveneens dat kleine verschillen in de (residuele) tussen-schoolse variantie tussen de te vergelijken studies of landen onterecht als relevant worden aangeduid. De onzekerheid omtrent de schattingen van de variantiecomponenten wordt bij het berekenen van de intra-klasse correlatie achterwege gelaten en ook bij de presentatie van intra-klasse correlaties is hier veelal geen aandacht voor (zie ook Figuur 1 van de discussiebijdrage). Er zijn inmiddels verschillende methoden ontwikkeld om betrouwbaarheids-intervallen rondom intra-klasse correlaties te berekenen, maar deze zijn vele malen complexer en zeker nog niet zonder problemen (e.g., Demetrashvili et al., 2016; Ionan et al., 2014).

4. Conclusie

Samenvattend heb ik in deze reactie proberen duidelijk te maken dat de genoemde problemen uit de discussiebijdrage van Lek et al. (2020) niet zozeer te maken hebben met intra-klasse correlatie zelf, maar dat de aangehaalde problematiek voortkomt uit de meerniveau-modellen die aan intra-klasse correlaties ten grondslag liggen. De suggesties die in de discussiebijdrage (Lek et al., 2020) worden gegeven zijn derhalve ook manieren om de onderliggende modellen te verbeteren. Deze suggesties voorkomen niet dat intra-klasse correlaties door de lezers van onderzoeksverslagen nog niet altijd op correcte wijze worden geïnterpreteerd. Om dubieuze interpretaties van intra-klasse correlaties in de toekomst te voorkomen, wil ik de auteurs van onderzoeksverslagen vragen veel explicieter te zijn dan tot nu toe gebruikelijk in de specificatie van het onderliggende meerniveau-model. Het is belangrijk om enerzijds duidelijk te maken naar welk hiërarchisch niveau de intra-klasse correlatie verwijst en anderzijds naar of correctiefactoren in het onderliggende model zijn toegepast en welke dat zijn. Samen dienen ze zicht te geven op de interpretatie die juist wel of niet aan de intra-klasse correlatie gegeven kan worden.

Noten

- ¹ In de discussiebijdrage wordt op pagina 326 het volgende gesteld: “waar $\sigma_{\text{leerling}}^2$ de variantie is in leerlingresultaten en $\sigma_{\text{leerling}}^2$ de variantie in gemiddelde leerlingresultaten over scholen”. Een correcte beschrijving van de variantiecomponenten is: waar $\sigma_{\text{leerling}}^2$ de variantie is in leerlingresultaten binnen scholen en σ_{school}^2 de variantie in gemiddelde leerlingresultaten over scholen
- ² Op pagina 326 van de discussie bijdrage wordt gesteld: “Heeft de ene school relatief veel leerlingen met hoge resultaten en de andere school relatief veel leerlingen met lage resultaten, dan neemt $\sigma_{\text{leerling}}^2$ relatief toe en gaat de intra-klasse correlatie omhoog.” Dit moet zijn: Heeft de ene school relatief veel leerlingen met

hoge resultaten en de andere school relatief veel leerlingen met lage resultaten, dan neemt σ^2_{school} relatief toe en gaat de intra-klasse correlatie omhoog.

Literatuur

- Bosker, R. J., Béguin, A., & Rekers-Mombarg, L. (2001). Hoe meten we de prestatie van een school? In A.B.Dijkstra, S. Karsten, R. Veenstra, & A. J. Visscher (Eds.), *Het oog der natie: Scholen op rapport* (pp. 121-135). Assen: Koninklijke Van Gorcum BV.
- Darmawan, I. G. N., & Keeves, J. P. (2006). Accountability of teachers and schools: A value-added approach. *International Education Journal*, 7, 174–188.
- De Fraine, B., Van Damme, J., & Onghena, P. (2002). Accountability of schools and teachers: What should be taken into account? *European Educational Research Journal*, 1, 403–428. <https://doi.org/10.2304/eeerj.2002.1.3.2>
- Demetrashvili, N., Wit, E. C., & Heuvel, van den, E. R. (2016). Confidence intervals for intraclass correlation coefficients in variance components models. *Statistical Methods in Medical Research*, 25(5), 2359–2376. <https://doi.org/10.1177/0962280214522787>
- Goldstein, H. (2001). Using pupil performance data for judging schools and teachers: scope and limitations. *British Educational Research Journal*, 27(4), 433–442.
- Goldstein, H., Browne, W. J., & Rasbash, J. (2002). *Partitioning variance in multilevel models*. London: Institute of Education.
- Goldstein, H., & Sammons, P. (1997). The influence of secondary and junior schools on sixteen year examination performance: A cross-classified multilevel analysis. *School Effectiveness and School Improvement*, 8, 219–230. <https://doi.org/10.1080/0924345970080203>
- Hill, P. W., & Goldstein, H. (1998). Multilevel modeling of educational data with cross-classification and missing identification for units. *Journal of Educational and Behavioral Parameters*, 23, 117–128. <https://doi.org/10.2307/1165317>
- Hill, P.W., & Rowe, K.J. (1996). Multilevel modelling in school effectiveness research. *School Effectiveness and School Improvement*, 7, 1–34
- Hox, J. (2010). *Multilevel analysis: Techniques and applications. Second Edition*. New York: Routledge.
- Inspectie van het Onderwijs. (2017). *De Staat van het Onderwijs 2015/2016*. Utrecht: Inspectie van het Onderwijs.
- Ionan, A. C., Polley, M.-Y. C., McShane, L. M., & Dobbin, K. K. (2014). Comparison of confidence interval methods for an intra-class correlation coefficient (ICC). *BMC Medical Research Methodology* 14, 121. <https://doi.org/10.1186/1471-2288-14-121>
- Lauder, H., Kounali, D., Robinson, T., & Goldstein, H. (2010). Pupil composition and accountability: An analysis of English primary schools. *International Journal of Educational Research*, 49, 49–68. <https://doi.org/10.1016/j.ijer.2010.08.001>
- Leckie, G. (2013). *Three-level multilevel models: Concepts. LEMMA VLE Module 11* (Bristol, Centre for Multilevel Modelling, University of Bristol).
- Leckie, G. & Goldstein, H. (2009). The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement. *Journal of the Royal Parametrical Society, Series A*, 172, 537–554. <https://doi.org/10.1111/j.1467-985X.2008.00577.x>
- Lek, K. Feskens, R. C. W., & Scheerens, J. (2020). Discussiebijdrage Schoolverschillen en schooleffecten in het voorgezet onderwijs. *Pedagogische Studiën*, 97, 324–335.
- McKinsey & Company (2020). *Een verstevigd fundament voor iedereen. Een onderzoek naar de doelmatigheid en toereikendheid van het funderend onderwijs (primair en voortgezet)*. Amsterdam: McKinsey & Company Netherlands.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 International Results in Reading*. International Association for the Evaluation of Educational Achievement (IEA).
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. International Association for the Evaluation of Educational Achievement (IEA).
- OECD (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. PISA, OECD Publishing, Paris.
- OECD (2019). *PISA 2018 Results (Volume II):*

- Where All Students Can Succeed. PISA, OECD Publishing, Paris,
- Peng, P., & Klieme, E. (2014). One-year value-added school effects from various models and their inter-temporal variability: Evidence from China. Unpublished manuscript. Retrieved from www.invalsi.it/invalsi/ri/improving_education/Papers/peng/151.doc
- Rasbash, J. & Browne, W. J. (2001). Modelling Non-Hierarchical Structures. In Leyland, A. H. and Goldstein, H. (Eds.), *Multilevel Modelling of Health Parameters*, 93-105. Chichester: John Wiley & Sons.
- Raudenbush, S. W. & Bryk, A. S. (1986). A Hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis* (2nd ed.). Thousand Oaks, CA: Sage.
- Scheerens J., & Bosker, R. J. (1997). Size, stability and consistency of school effects. In: J. Scheerens & R.J. Bosker, *The foundations of educational effectiveness* (pp. 69-96). Oxford: Pergamon.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modelling* (2nd ed.). Los Angeles, CA: Sage.
- Tekwe, C. D., & Carter, R. L., Ma, C.-X., Algina, J., Lucas, M. E., Roth, J., ... Resnick, M. B. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics*, 29, 11–36. <https://doi.org/10.3102/10769986029001011>
- Thomas, S., & Mortimore, P. (1996). Comparison of value-added models for secondary-school effectiveness. *Research Papers in Education*, 11, 5–33. <https://doi.org/10.1080/0267152960110103>
- Timmermans, A. C. (2012). *Value added in educational accountability: possible, fair and useful?* GION, Gronings Instituut voor Onderzoek van Onderwijs, Rijksuniversiteit Groningen.
- Timmermans, A. C., Doolaard, S., & De Wolf, I. (2011). Conceptual and empirical differences among various value-added models for accountability. *School Effectiveness and School Improvement*, 22, 393–413. <https://doi.org/10.1080/09243453.2011.590704>
- Timmermans, A. C., Snijders, T. A. B., & Bosker, R. J. (2013). In Search of Value Added in the Case of Complex School Effects. *Educational and Psychological Measurement*, 73(2), 210-228
- Timmermans, A. C., & Thomas, S. (2015). The Impact of Student Composition on Schools Value Added Performance: A Comparison of Seven Empirical Studies. *School Effectiveness and School Improvement*. 26 (3), 487 – 498. <https://doi.org/10.1080/09243453.2014.957328>
- Timmermans, A.C. & Van der Werf, G. (2017). Accounting for previous performance of students by means of growth curves analyses to estimate the size, stability and consistency of school effects. *Educational Research and Evaluation*, 23(5-6), 221 -246.

Auteur

Anneke Timmermans is universitair hoofddocent aan de Rijksuniversiteit Groningen, Faculteit Gedrags- en Maatschappijwetenschappen, GION Onderwijs/Onderzoek.

Correspondentieadres: A. C. Timmermans, Rijksuniversiteit Groningen, GION Onderwijs/Onderzoek, Grote Rozenstraat 3, 9712 TG Groningen. E-mail: A.C.Timmermans@rug.nl