



EVALUATIE & DIVERSITEIT IN HET BASIS- EN SECUNDAIR ONDERWIJS

Deelrapport: survey-onderzoek

Joke Ysenbaert, Piet Van Avermaet & Mieke Van Houtte



EVALUATIE & DIVERSITEIT IN HET BASIS- EN SECUNDAIR ONDERWIJS

Deelrapport: survey-onderzoek

Joke Ysenbaert, Piet Van Avermaet & Mieke Van Houtte

Promotor: Prof. Dr. Piet Van Avermaet

**Copromotoren: Prof. Dr. Kris Van den Branden & Prof. Dr. Mieke
Van Houtte**

Research paper SONO/2020.OL1.6/1

Gent, maart 2020

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



Gelieve naar deze publicatie te verwijzen als volgt:

Ysenbaert, J., Van Avermaet, P., & Van Houtte, M. (2020). Evaluatie & diversiteit in het basis- en secundair onderwijs: Eindrapport surveyonderzoek.

Voor meer informatie over deze publicatie Joke.Ysenbaert@UGent.be

Deze publicatie kwam tot stand met de steun van de Vlaamse Gemeenschap, Ministerie voor Onderwijs en Vorming.

In deze publicatie wordt de mening van de auteur weergegeven en niet die van de Vlaamse overheid. De Vlaamse overheid is niet aansprakelijk voor het gebruik dat kan worden gemaakt van de opgenomen gegevens.

© 2020 STEUNPUNT ONDERWIJSONDERZOEK

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

Deze publicatie is ook beschikbaar via www.steunpuntsono.be

Voorwoord

Dit rapport is een onderdeel van het onderzoek dat wordt uitgevoerd binnen het Steunpunt voor Onderwijsonderzoek (SONO). SONO heeft als doel om gedurende de periode 2016-2020 beleidsvoorbereidend onderzoek te voeren rond verschillende thema's die door de Vlaamse Overheid als prioritair en relevant beschouwd worden. Eén van de onderzoekslijnen (OL. 1.6) focust op het evaluatiebeleid van scholen en de evaluatiepraktijken van leerkrachten in het basis- en secundair onderwijs. Binnen deze onderzoekslijn werden reeds twee rapporten gepubliceerd: de literatuurstudie¹ en het rapport met de resultaten van het casestudieonderzoek². Dit kwalitatieve onderzoeksluik wordt aangevuld met een kwantitatief luik waar dit rapport de neerslag van vormt. In een volgende fase zullen de resultaten en bevindingen van de drie deelstudies gebundeld worden in één eindrapport.

Het voorliggend document rapporteert over het grootschalig surveyonderzoek dat uitgevoerd werd in 2019 in zowel het basis- als het secundair onderwijs. In het eerste hoofdstuk wordt gekaderd waarom de evaluatiepraktijk van leerkrachten en het evaluatiebeleid van scholen onderzocht wordt met specifieke aandacht voor de rol die diversiteit speelt. De focus ligt niet louter op leerlingenevaluatie op zich, maar ook in hoe dit zich verhoudt tot de mate waarin men in de algemene klaspraktijk aan differentiatie doet, de opvattingen van leerkrachten over de rol van onderwijs ten aanzien van diversiteit en de mate waarin men zich bekwaam voelt om diversiteit onder leerlingen op te merken.

Het tweede hoofdstuk licht het theoretisch kader toe dat bepalend is voor welke variabelen op school- en leerkrachtniveau mee opgenomen werden in de vragenlijsten. De ontwikkeling van de vragenlijsten, de steekproeftrekking en het verloop van de dataverzameling wordt beschreven in het derde hoofdstuk. Het vierde en vijfde hoofdstuk lichten ten slotte de analyses en resultaten toe van het onderzoek. Eerst komen de bevindingen van de beschrijvende univariate analyses (hoofdstuk 4) aan bod. De verklarende multivariate analyses worden beschreven in het vijfde hoofdstuk.

Voorafgaand aan de verschillende hoofdstukken, geven we eerst een beleidssamenvatting.

¹ Ysenbaert, J., Van Avermaet, P., & Van Houtte, M. (2017). Literatuurstudie evaluatie en diversiteit, Steunpunt Onderwijsonderzoek, Gent.

² Ysenbaert, J., Van Avermaet, P., & Van Houtte, M. (2018). *Evaluatie & diversiteit in het basis- en secundair onderwijs. Deelrapport: casestudieonderzoek*, Steunpunt Onderwijsonderzoek, Gent.

Beleidssamenvatting

Het hoeft geen betoog meer dat veel leerkrachten vandaag de dag uitgedaagd worden door de diversiteit onder leerlingen in de klas. Van leerkrachten wordt verwacht dat ze op een inclusieve manier tegemoet komen aan de leernoden van álle leerlingen in een klas, of ze nu onderling verschillen met betrekking tot de gevolgde schoolloopbaan, schoolse prestaties, achtergrond, interesse, thuistaal, motivatie, ... In navolging van deze realiteit in het werkveld, hanteren we in dit deelrapport de inclusieve benadering van diversiteit. Diversiteit wordt aldus breed en intersectorieel gedefinieerd, dit in tegenstelling tot een enge, gefragmenteerde benadering waarbij gefocust wordt op een specifieke doelgroep die afgebakend wordt op basis van één bepaald kenmerk, bijvoorbeeld etniciteit.

Deze diversiteit stelt niet alleen uitdagingen voor de lespraktijk, maar aangezien kwaliteitsvol onderwijs vraagt dat evaluatiepraktijken afgestemd zijn op de lespraktijk, brengt dit ook uitdagingen met zich mee voor de evaluatiepraktijk van leerkrachten en het evaluatiebeleid van scholen.

Het voorwerp van dit survey-onderzoek vormt dan ook de evaluatiepraktijk van leerkrachten. We willen nagaan hoe evaluatiepraktijken eruit zien en vorm krijgen naarmate de context gekenmerkt wordt door diversiteit onder leerlingen. Om dit te onderzoeken maken we gebruik van de eerder ontwikkelde modellen voor evaluatiepraktijk³: het gescheiden model, het semi-geïntegreerde en geïntegreerde model. Gebaseerd op het model van Kozma en Voogt (2003), dat een onderscheid maakt tussen factoren op micro- of leerkrachtniveau en factoren op meso- of schoolniveau, maken we gebruik van multilevelanalyses om na te gaan welke factoren een rol spelen.

Via een online survey werden zowel leerkrachten als directies bevestigd met betrekking tot het evaluatiebeleid van de school en hun evaluatiepraktijken. De survey peilde naar hoe hun evaluatiepraktijk zich verhoudt tot hun algemene lespraktijk, in welke mate men bij evaluatie aandacht heeft voor aspecten zoals eerlijkheid, transparantie, verantwoording, impact en betrokkenheid van leerlingen. De focus van dit onderzoek ligt niet alleen op leerlingenevaluatie in het algemeen, maar ook op de mate waarin hierbij rekening gehouden wordt met diversiteit onder leerlingen. Hiertoe werd naast de opvattingen omtrent evaluatie ook gepeild naar de opvattingen ten aanzien van de rol van onderwijs inzake het omgaan met diversiteit onder leerlingen en naar de mate waarin leerkrachten zichzelf bekwaam voelen om diversiteit onder leerlingen op te merken.

Voor dit onderzoek werden twee vragenlijsten ontwikkeld: één voor directies en één voor leerkrachten. Deze werden afgenomen in 27 basisscholen en 29 secundaire scholen. Zowel scholen met veel diversiteit in het leerlingenpubliek als scholen met een eerder homogeen

³ Ysenbaert, J., Van Avermaet, P., & Van Houtte, M. (2017). Literatuurstudie evaluatie en diversiteit, Steunpunt Onderwijsonderzoek, Gent.

leerlingenpubliek werden meegenomen in het onderzoek om te kijken of er met betrekking tot leerlingencompositie schoolspecifieke tendensen vastgesteld kunnen worden op vlak van evaluatiebeleid- en praktijk. In totaal werden 47 directeurs en 1120 leerkrachten bereikt die deelnamen aan het onderzoek.

Voor de bevraging bij leerkrachten werd gebruik gemaakt van drie modellen die representeren welke plaats evaluatie krijgt in de ruimere onderwijspraktijk. Het **gescheiden model**, waarbij de instructieactiviteiten sterk gescheiden zijn van de evaluatieactiviteiten, komt over het algemeen het frequentst voor. Zowel leerkrachten van het basis- als van het secundair onderwijs geven aan dat dit model vaak kenmerkend is voor hun praktijk. Leerkrachten die inzetten op het gescheiden model lijken in hun lespraktijk over het algemeen weinig in te zetten op differentiatie en op het betrekken van leerlingen tijdens het evaluatieproces. Bovendien besteden ze veel aandacht aan het kunnen verantwoorden van de uitkomsten van de evaluatie.

Bij het **semi-geïntegreerde model** haken instructieactiviteiten en evaluatieactiviteiten meer op elkaar in: evaluatie gaat aan de instructie vooraf waardoor deze laatste afgestemd wordt op wat uit de evaluatie voortvloeit. Kenmerkend voor dit model is dat leerkrachten wiens evaluatiepraktijk hierbij aansluit ook in het algemeen meer aandacht hebben voor differentiatie in hun lessen. Bovendien zullen ze ook meer geneigd zijn om leerlingen te betrekken bij het evaluatieproces en zorgen ze ervoor dat evaluatie impact kan hebben op het uiteindelijke leren van leerlingen. Dit semi-geïntegreerde model is in mindere mate representatief voor de evaluatiepraktijk van de bevroegde leerkrachten. Leerkrachten sluiten met hun praktijk minder vaak aan bij dit model dan bij het gescheiden model. Hoewel weinig verschil, stellen we vast dat het basisonderwijs hier gemiddeld iets hoger scoort dan het secundair onderwijs.

Het **geïntegreerde model**, ten slotte, verwijst naar een sterke inbedding van evaluatie in de ruimere lespraktijk. Een evaluatiepraktijk die gekenmerkt wordt door een sterke integratie van evaluatie en instructie gaat samen met aandacht voor differentiatie in het algemeen en met het nauw betrekken van leerlingen tijdens het evaluatieproces. In vergelijking met de twee vorige modellen is dit het minste herkenbaar in de evaluatiepraktijk van de respondenten: leerkrachten geven aan dat dit soms tot eerder weinig aanwezig is in hun praktijk. Net zoals bij het semi-geïntegreerde model komt dit geïntegreerde model nog iets vaker voor in het basisonderwijs dan in het secundair onderwijs.

Géén van deze modellen kan beschouwd worden als een typerend kenmerk van de school. Dit betekent dat de evaluatiepraktijken van leerkrachten onderling sterk verschillen en dus beperkt gedeeld worden binnen eenzelfde school. We kunnen scholen dus niet van elkaar onderscheiden op basis van (modellen van) hun evaluatiepraktijken. We stelden vast dat alle leerkrachten, ongeacht het model waarbij ze aansluiten, aandacht besteden aan eerlijke en transparante evaluatie.

De multilevelanalyses laten zien dat enkele factoren op schoolniveau bepalend zijn voor de mate waarin leerkrachten al dan niet zullen aansluiten bij één van de drie modellen. Zo blijkt de frequentie waarmee leerlingcontacten of gesprekken tussen leerkracht en leerling over het leerproces georganiseerd worden positief samen te hangen met de mate waarin leerkrachten hun evaluatiepraktijk afstemmen op het gescheiden model. Er is een positieve samenhang van zodra er meer dan twee leerlingcontacten per schooljaar georganiseerd worden. Aangezien dit model ook getypeerd wordt door een sterke aandacht voor het verantwoordingsaspect, verwachten we dat

de leerlingcontacten eerder hierop de nadruk leggen. Een tweede factor op schoolniveau blijkt het studieaanbod te zijn: leerkrachten op scholen met een studieaanbod bso zijn minder geneigd om beroep te doen op het gescheiden model. Leerkrachten die in het bso een lesopdracht hebben, blijken veel meer aan te sluiten bij het geïntegreerde model. Dit kan verklaard worden door enerzijds de praktijkgerichte vakken die typerend zijn voor deze onderwijsvorm: vaardigheden staan hier centraal en de evaluatie/bijsturing van deze vaardigheden vindt meer plaats tijdens het lesgebeuren op het moment dat de leerlingen ze uitoefenen en ontwikkelen. Anderzijds blijkt de leerlingencompositie van de school een bepalende factor te zijn voor de evaluatiepraktijk van leerkrachten: hoe meer het leerlingenpubliek van een school gekenmerkt wordt door een hoge SES, hoe minder de leerkrachten zullen kiezen voor een (semi-)geïntegreerde evaluatiepraktijk. Leerkrachten zullen hun evaluatiepraktijk meer integreren in de algemene lespraktijk als ze les geven op een school die gekenmerkt wordt door een groter aandeel lage SES-leerlingen. Aangezien het bso over het algemeen meer leerlingen met een lagere SES heeft, bevestigt dit de bevinding dat leerkrachten in het bso vaker de evaluatie en instructie op elkaar laten aansluiten. Ten slotte blijkt een sterker beleidsvoerend vermogen van de school inzake het evaluatiebeleid positief samen te hangen met de mate waarin leerkrachten zullen kiezen voor een geïntegreerde evaluatiepraktijk.

Naast de factoren op schoolniveau blijken ook factoren op het niveau van de individuele leerkracht bepalend te zijn voor de mate waarin ze aansluiten bij een gescheiden of (semi-)geïntegreerd model. Zo blijken de opvattingen van leerkrachten een rol te spelen: leerkrachten die overtuigd zijn van de informerende waarde van evaluatie zullen in hun praktijk aansluiten bij het gescheiden model, leerkrachten die eerder kritisch zijn ten aanzien van evaluatie omdat ze menen dat evaluatie meetfouten kan bevatten en oneerlijk kan zijn voor sommige leerlingen, zullen meer het geïntegreerde model toepassen. Verder blijkt ook dat de mate waarin leerkrachten inzetten op professionalisering eerder samenhangt met (semi-)geïntegreerde evaluatiepraktijken. Tegelijk blijkt ook dat die leerkrachten die zichzelf bekwaam voelen om diversiteit onder leerlingen op te merken vaker het (semi-)geïntegreerde model zullen aanwenden. Mogelijks gaan deze leerkrachten specifiek op zoek naar professionalisering, precies omdat zij met de diversiteit onder leerlingen ook aan de slag willen gaan. Leerkrachten die ervan overtuigd zijn dat wat ze doen in de praktijk ook een positief effect heeft op het leren en ontwikkelen van leerlingen, zullen ook vaker aansluiten bij het geïntegreerde model. Ten slotte bleek ook dat het geslacht van de leerkracht een positief verband heeft met het geïntegreerde model: mannen zouden eerder dan vrouwen gebruik maken van een evaluatie die sterk ingebed is in het ruimere onderwijsleerproces. Dit verband houdt bovendien stand na controle voor onderwijsvorm. Aangezien dit betekent dat de leerlingen zelf heel sterk betrokken worden, sluit dit positieve effect van 'geslacht' aan bij eerder onderzoek dat aantoont dat mannen meer dan vrouwen inzetten op het ondersteunen van autonomie (Lietaert & Consuegra, 2015).

Inhoud

Voorwoord	1
Beleidssamenvatting	2
Inhoud	5
Hoofdstuk I Achtergrond	7
1. Evaluatiebeleid in een context van diversiteit	7
2. Bekeken vanuit een multilevel benadering	8
Hoofdstuk II Theoretisch kader	9
1. Evaluatiebeleid & -praktijk als voorwerp van onderzoek	9
1.1. Evaluatiepraktijk	10
1.2. Opvattingen met betrekking tot evaluatie	13
2. Factoren op leerkrachtniveau	14
2.1. Opvattingen & attitudes leerkrachten	14
2.2. Bekwaamheid & doelmatigheidsbeleving	15
2.3. Ervaring	16
3. Factoren op schoolniveau	16
3.1. Beleidsvoerend vermogen	17
3.2. Studieaanbod	18
3.3. Leerlingencompositie	19
Hoofdstuk III Methodologie	20
1. Onderzoeksopzet & onderzoeksvragen	20
2. Steekproef	25
2.1. Basisonderwijs	25
2.2. Secundair onderwijs	26
3. Instrumenten	31
3.1. Vragenlijst leerkrachten	31
3.2. Vragenlijst directies	31
3.3. Ontwikkeling instrumenten	32
4. Verloop dataverzameling	33
4.1. Opvolging dataverzameling	33
4.2. Respons	33
Hoofdstuk IV Univariate analyses	34
1. Beschrijving respondenten	34
1.1. Leerkrachten	34
1.2. Directies	39
2. Beschrijving procesvariabelen	41
2.1. Differentiatiepraktijk leerkrachten	41
2.2. Aandacht voor de kwaliteitscriteria impact, betrokkenheid, eerlijkheid, transparantie en verantwoording	41
2.3. Professionalisering van leerkrachten	43
2.4. Opvattingen van leerkrachten met betrekking tot onderwijs	43
2.5. Opvattingen van leerkrachten met betrekking tot diversiteit	44
2.6. Bekwaamheidsgevoel van leerkrachten met betrekking tot oog hebben voor diversiteit	44
2.7. Opvattingen van leerkrachten met betrekking tot evaluatie	44

2.8. Doelmatigheidsbeleving leerkrachten	46
2.9. Organisatie van de school	46
2.10. Beleidsvoerend vermogen van de school inzake evaluatie	49
3. Beschrijving output-variabelen	50
Hoofdstuk V Multivariate analyses	52
1. Werkwijze	52
2. Modellen voor evaluatiepraktijk	58
2.1. Gescheiden model als afhankelijke variabele	59
2.2. Geïntegreerd model als afhankelijke	60
2.3. Semi-geïntegreerd model als afhankelijke	60
2.4. Conclusie met betrekking tot modellen voor evaluatiepraktijk	61
3. Multilevelanalyses	62
3.1. Methode	62
3.2. Modelopbouw	64
3.3. Resultaten	64
3.4. Conclusies bij multilevelanalyses voor de evaluatiepraktijk van leerkrachten.	78
Bibliografie	84

Hoofdstuk I Achtergrond

1. Evaluatiebeleid in een context van diversiteit

Het hoeft geen betoog meer dat veel leerkrachten vandaag de dag uitgedaagd worden door de diversiteit onder leerlingen in de klas. Deze diversiteit stelt niet alleen uitdagingen voor de lespraktijk, maar aangezien kwaliteitsvol onderwijs vraagt dat evaluatiepraktijken afgestemd zijn op de lespraktijk, brengt dit ook uitdagingen met zich mee voor de evaluatiepraktijk van leerkrachten en bijgevolg ook voor het evaluatiebeleid van scholen.

De laatste decennia zien we dat er wereldwijd beleidsinitiatieven genomen worden die inzetten op het inrichten van leeromgevingen die aandacht hebben voor diversiteit (UNESCO, 1994). Specifiek voor de context van Vlaanderen zien we dit vertaald in onder andere het GOK-beleid en, meer recent, de invoering van het M-decreet. Leerkrachten, van lager tot secundair onderwijs hebben de taak om in te spelen op de verschillen tussen leerlingen en zo elke lerende te ondersteunen in het bereiken van de leerdoelen (Ainscow, 2007; Banks & al., 2007; Van Avermaet, Van Houtte, & Van den Branden, 2011). In de literatuur wordt hiervoor vaak verwezen naar 'gedifferentieerde instructie' (C.A. Tomlinson & Moon, 2013) als manier om onderwijs niet langer als 'one-size-fits-all' in te richten, maar als manier om tegemoet te komen aan de noden van leerlingen om kwaliteitsvol onderwijs voor elke leerling te garanderen (Coubergs, Struyven, Vanthournout, & Engels, 2017). Uiteraard vergt dit niet alleen wijzigingen in de lespraktijk, maar eveneens in de evaluatiepraktijk van leerkrachten (T. R. Moon, 2005; Tonya R Moon, 2016; Carol Ann Tomlinson, 2005). Immers, in de literatuur wordt benadrukt dat afstemming tussen lespraktijk en evaluatiepraktijk noodzakelijk zijn voor kwaliteitsvol onderwijs (James, 2006; Shepard, 2000; Tyler, 2013). Onderwijs dat rekening houdt met verschillen tussen leerlingen, vraagt aldus om evaluatiepraktijken die ook rekening houden met deze verschillen onder leerlingen.

Met betrekking tot evaluatie zien we dat er in de literatuur een paradigmaverschuiving heeft plaats gevonden, meer specifiek heeft de klassieke *toetscultuur* plaats gemaakt voor een *assessmentcultuur* (Birenbaum, 2014; Dierick & Dochy, 2001; Falchikov, 2013). Beide worden vaak zo gepresenteerd alsof het tegenpolen zijn. In de literatuur zien we dat beide strekkingen uitgegroeid zijn tot paradigma's die één specifieke functie van evaluatie belichten, vaak ten koste van de andere functie. Zo ligt binnen de testcultuur de nadruk op het beoordelen of een doel al dan niet bereikt is. Dit is in tegenstelling tot de *assessmentcultuur* waarbij evaluatie veel meer gezien wordt als een manier om het onderwijsleerproces te ondersteunen. De verschuiving naar een *assessmentcultuur* heeft ook gevolgen voor hoe 'kwaliteitsvolle evaluatie' ingevuld wordt (Segers & Tillema, 2011). Binnen de toetscultuur ligt de nadruk veel meer op *psychometrische* criteria waarbij objectieve en gestandaardiseerde evaluaties als meer kwaliteitsvol gezien worden. De *assessmentcultuur* legt daarentegen meer de nadruk op *edumetrische* criteria waarbij evaluatie de mogelijkheid moet bieden aan de lerende om eigenaar te worden van het eigen leerproces. Hiervoor worden specifieke vormen van *assessment* zoals zelf- en peerevaluatie naar voren geschoven omdat ze de mogelijkheid bieden om in dialoog en interactie te gaan. Hoewel beide paradigma's vaak als tegenpolen geschetst worden, stellen we vast dat ze beide waardevol zijn voor de klaspraktijk. Ze stellen elk een specifieke functie

van evaluatie centraal. Functies die noodzakelijk zijn in de dagdagelijkse onderwijspraktijk, waardoor beide paradigma's – hoewel ze soms in spanning staan tot elkaar - betekenisvol zijn binnen de actuele onderwijspraktijk.

Met betrekking tot 'diversiteit' zien we dat beide paradigma's een verschillend perspectief hanteren. De *toetscultuur* gaat uit van de gedachte dat het mogelijk is om de bekwaamheid van een leerling te capteren in één moment, tijdens de afname van een toets. Omwille van de objectiviteit ligt de nadruk eveneens op 'iedereen wordt gelijk geëvalueerd'. De *assessmentcultuur* daarentegen vertrekt veel meer vanuit het feit dat het leren voortdurend in ontwikkeling is en niet zozeer meetbaar is via één momentopname. De nadruk ligt hier immers veel meer op het proces van deze ontwikkeling, dan op de uitkomst ervan. Aangezien evaluatie gezien wordt als manier om het onderwijsleerproces te ondersteunen, wordt diversiteit onder leerlingen veel meer erkend binnen de assessmentcultuur, precies omdat het leerproces van leerlingen verschillend kan verlopen.

2. Bekeken vanuit een multilevel benadering

Het voorwerp van het survey-onderzoek vormt het evaluatiebeleid van scholen en de evaluatiepraktijk van leerkrachten. Concreet zullen we nagaan hoe evaluatiepraktijken eruit zien en vorm krijgen naarmate de context gekenmerkt wordt door diversiteit onder leerlingen. Om dit te onderzoeken doen we beroep op het model van Kozma and Voogt (2003). Hoewel de auteurs zich toespitsen op onderwijstechnologie en hoe individuele en contextuele factoren het gebruik hiervan in de klaspraktijk beïnvloeden, kan dit model ook als raamwerk fungeren om de evaluatiepraktijken van leerkrachten te verklaren. Volgens dit raamwerk worden drie niveaus onderscheiden: micro-, meso- en macroniveau. Deze drie niveaus verwijzen naar factoren die gerelateerd zijn aan de leerkracht, de school en het ruimere onderwijssysteem die een invloed uitoefenen op de evaluatiepraktijk (Black & Wiliam, 2005; Bonner, 2016).

Het microniveau verwijst specifiek naar de kenmerken en de praktijken van wat op klasniveau en dus ook bij de individuele leerkracht gebeurt. Het mesoniveau verwijst naar deze factoren die zich buiten de klas bevinden, maar mogelijks een effect in de klas teweeg brengen. Dit kan bijvoorbeeld ondersteuning van het schoolbeleid zijn of de organisatie van leergemeenschappen binnen de school zelf. Het macroniveau verwijst dan weer naar factoren die verder afstaan van de school zelf, maar er wel een invloed op uitoefenen. In het kader van evaluatie denken we hier bijvoorbeeld aan het wettelijk kader dat stelt dat scholen over veel vrijheid beschikken om een eigen evaluatiebeleid te ontwikkelen (Penninckx, Vanhoof, & Van Petegem, 2011).

De literatuur inzake schooleffectiviteitsonderzoek stelt bovendien dat deze drie niveaus op elkaar inwerken: factoren op het mesoniveau kunnen processen op het microniveau faciliteren of net belemmeren, maar ook omgekeerd (Kozma & Voogt, 2003; Scheerens, 1990).

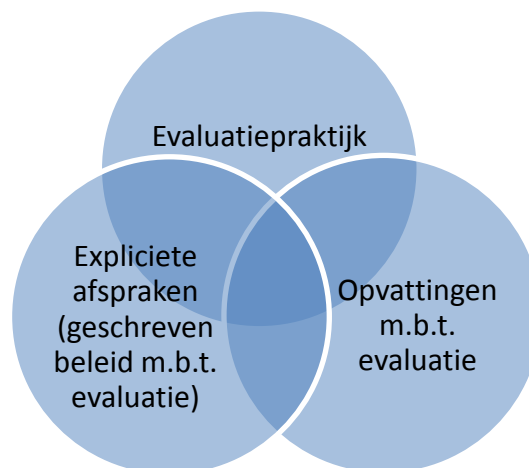
Hoofdstuk II Theoretisch kader

1. Evaluatiebeleid & -praktijk als voorwerp van onderzoek

Binnen dit onderzoek beschouwen we het evaluatiebeleid van een school als het samengaan van drie componenten: evaluatiepraktijken van leerkrachten, opvattingen omtrent evaluatie en expliciete afspraken op schoolniveau omtrent evaluatie. Deze driedeling is gebaseerd op Spolsky's benadering van 'language policy':

'language policy exists even where it has not been made explicit or established by authority' (Spolsky, 2004).

Deze benadering van het concept 'beleid' zal als blauwdruk fungeren om het 'evaluatiebeleid' van scholen in Vlaanderen te onderzoeken. Het evaluatiebeleid kan best onderzocht worden door aandacht te hebben voor het samengaan van deze drie componenten. Leerkrachten en directies houden er immers opvattingen op na omtrent hoe evaluatie zou moeten zijn. Deze opvattingen oefenen ook een invloed uit op de evaluatiepraktijk zelf. De evaluatiepraktijk verwijst dan weer naar wat leerkrachten precies doen om hun leerlingen te evalueren. Het geëxpliciteerde evaluatiebeleid of de afspraken hieromtrent op schoolniveau kunnen deze opvattingen ondersteunen of net ingaan tegen de ideeën die leerkrachten erop nahouden. Het evaluatiebeleid van een school kan dus niet verengd worden tot louter het geëxpliciteerde beleid omtrent evaluatie, maar moet gezien worden als het samengaan van deze drie componenten die elkaar onderling beïnvloeden.



Onderzoek naar evaluatie in het Vlaamse secundair onderwijs toonde eerder al aan dat de meerderheid van de scholen niet over een expliciet, geschreven beleid omtrent evaluatie beschikt (Struyf, 2000). Tegelijk toont onderzoek van Verhoeven, Devos, Bruylant, and Warmoes (2002) aan dat zowel leerkrachten als directies ervan overtuigd zijn dat er een evaluatiebeleid is op school. Deze vaststellingen werden ook bevestigd in het casestudie-onderzoek dat aan dit survey-onderzoek vooraf ging. In de meerderheid van de twaalf onderzochte scholen beperkte het geschreven evaluatiebeleid

zich tot dat wat opgenomen was in het schoolreglement. Daarin lag de nadruk vooral op het meedelen van afspraken omtrent de frequentie van evaluaties (zoals het aantal examenperiodes) en het juridisch kader, zoals bijvoorbeeld de bevoegdheden van de klassenraad bij het toekennen van attesten (Ysenbaert, Van Avermaet, & Van Houtte, 2018). Omwille van de beperkte invulling van het expliciete, geschreven evaluatiebeleid van scholen, wordt dit niet verder opgenomen als focus in deze fase van het onderzoek. Dit survey-onderzoek wil de aandacht voornamelijk vestigen op de evaluatiepraktijk van leerkrachten en hun opvattingen met betrekking tot evaluatie.

1.1. Evaluatiepraktijk

Op basis van de literatuurstudie (Ysenbaert, Van Avermaet, & Van Houtte, 2017), die eerder uitgevoerd werd binnen dit project, maken we een onderscheid tussen drie types of modellen van evaluatiepraktijk: het gescheiden model, het semi-geïntegreerd model en het geïntegreerd model. Deze drie modellen zijn afgeleid van de drie belangrijkste paradigma's met betrekking tot leren en onderwijzen – behaviorisme, cognitivisme en constructivisme – elk met een eigen benadering van hoe leren ontstaat en hoe de onderwijspraktijk hierop kan inspelen. Net omdat evaluatie ingrijpt op dit onderwijsleerproces (Remesal, 2011) krijgt evaluatie, naargelang het paradigma dat gehanteerd wordt, een andere rol in dit onderwijsleerproces. In wat volgt, worden de drie modellen nog eens kort toegelicht.

In het gescheiden model zijn de instructieactiviteiten sterk gescheiden van de evaluatieactiviteiten: beide worden als aparte activiteiten beschouwd. Evaluatie wordt gezien als een manier om te controleren of de instructie succesvol was voor de leerling, waarbij minder goede uitkomsten al of niet resulteren in remediëring vooraleer over te gaan tot het aanleren van nieuwe leerstofgehelen. Hierbij wordt vooral de nadruk gelegd op evalueren van wat geleerd werd (Ertmer & Newby, 2013; Schunk, 2012; Shepard, 2000). Het vormt een soort van controle of de instructie haar doel bereikt heeft bij de lerende.



Fig. 1: Gescheiden model

Volgens het semi-geïntegreerde model vormen evaluatie en instructie geen losse, van elkaar staande activiteiten, maar grijpen ze op elkaar in. Binnen het cognitivisme wordt leren namelijk gezien als het 'ophangen' van nieuwe kennis aan reeds bestaande kennisschema's. Leren wordt geconcipieerd als het steeds verder uitbreiden van mentale structuren. Om dit leren te bevorderen wordt de voorkennis van leerlingen aangesproken om de nieuwe leerstof hierop te laten aansluiten. Het is net hierin dat evaluatie ook een rol kan spelen volgens het semi-geïntegreerde model: via evaluatie, bijvoorbeeld in de vorm van vraagstelling bij aanvang van de les, kan de voorkennis van leerlingen geactiveerd worden. Afhankelijk van de mate waarin de lerende reeds over voorkennis beschikt, kan de leerkracht zijn/haar instructieactiviteiten hier op afstemmen. Via dit inhaken van instructie op de uitkomsten van evaluatie, wordt voorkomen dat nieuwe kennis te ver zou afstaan van reeds bestaande mentale structuren (James, 2006). Dit model benadrukt veel meer de functie van 'evalueren om te leren', dit in tegenstelling tot het gescheiden model dat de functie van 'evalueren van het leren' benadrukt.

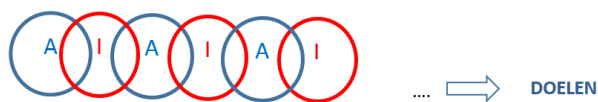


Fig. 2: Semi-geïntegreerd model

Het geïntegreerde model, dat afgeleid is van het constructivisme, beklemtoont veeleer de functie van 'evalueren als leren'. In dit model is er geen helder onderscheid meer tussen de activiteit van leren en van evalueren: beide lijken haast samen te vallen. Evalueren wordt namelijk gezien als leren.

Dit model benadrukt sterk de verantwoordelijkheid die de lerende heeft in zijn/haar eigen leerproces. De leerkracht daarentegen treedt hier – in vergelijking met de andere modellen – minder op de voorgrond en heeft meer een begeleidende rol. 'Evaluatie' is bijgevolg niet langer een activiteit die vooral vertrekt vanuit de leerkracht, maar die eveneens kan vertrekken vanuit de lerende zelf en/of de *peers*. In dit model vallen evaluatie en instructie quasi samen omdat evaluatie gezien wordt als cruciaal element in het stimuleren van de metacognitie en de zelfregulerende vaardigheden die nodig zijn voor de evolutie in het leerproces (Clark, 2012; Dann, 2014; Earl, 2003).



Fig. 3: Geïntegreerd model

Net zoals de drie paradigma's op onderwijs en leren ontstaan zijn door wijzigende visies en opvattingen, is ook de wijze waarop evaluatie benaderd wordt geëvolueerd. Dit werd mede versterkt door veranderende verwachtingen van het onderwijs waarbij de nadruk meer kwam te liggen op het nastreven van competenties zoals bijvoorbeeld de 21^e eeuwse vaardigheden en competenties. Specifiek met betrekking tot evaluatie wordt deze paradigmaverschuiving omschreven als de evolutie van een toetscultuur naar een assessmentcultuur (Gipps, 2011). De toetscultuur legt de nadruk op het evalueren van het leren wat resulteert in het evalueren van een product waarbij er vooraf een norm bepaald wordt. Evaluatie is in dit geval sterk leerkrachtgestuurd waarbij het evaluatieproces er voor alle lerenden hetzelfde uitziet. De assessmentcultuur focust veeleer op de impact van evaluatie op het uiteindelijke leerproces. Evaluatie wordt in deze benadering 'breed' ingevuld: niet alleen de leerkracht evalueert, maar ook de leerling zelf en/of de medeleerlingen. Niet zozeer een vooraf bepaalde norm is hier het criterium, maar de leerling zelf vormt het vergelijkingspunt waardoor het geheel van instructie en evaluatie samenkomen in een model dat zich richt op het nastreven van groei. Aangezien de leerling een veel centralere plaats heeft binnen de assessmentcultuur, wordt in het evaluatieproces de diversiteit onder leerlingen meer erkend dan binnen een toetscultuur.

Met betrekking tot de eerder omschreven modellen voor evaluatiepraktijk kunnen we vaststellen dat het gescheiden model eerder aansluit bij de toetscultuur; het geïntegreerde model leunt veel meer aan bij de assessmentcultuur. Indien we deze twee modellen als uitersten van een continuüm zouden beschouwen, zouden we kunnen verwachten dat het semi-geïntegreerde model zich meer in het midden van dit continuüm situeert. Hier is de evaluatie in mindere mate gestuurd door de leerkracht in vergelijking met het gescheiden model en de leerling wordt meer betrokken, al is dit nog minder dan in vergelijking met het geïntegreerde model.


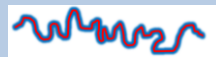
Toetscultuur	Assessmentcultuur
Focus op behalen van de norm	Focus op het leren
Evaluëren van het leren	Evaluëren om te/als leren
Leerkrachtgestuurde evaluatie	Leerling actief betrokken
Product	Product en proces
Statisch model	Groeimodel
Kennis	Competenties
<i>One-size-fits-all</i> -benadering	Erkent diversiteit onder leerlingen
	

Fig. 4: Kenmerkende verschillen tussen toetscultuur en assessmentcultuur

De verschuiving van een toetscultuur naar een bredere assessmentcultuur heeft veranderende inzichten met zich meegebracht met betrekking tot de criteria voor kwaliteitsvolle evaluatie. De klassieke evaluatiecriteria, zoals betrouwbaarheid en representativiteit, vormden niet langer dé enige kwaliteitscriteria voor degelijke evaluatie binnen de klaspraktijk (Birenbaum, 2016; Dierick & Dochy, 2001). In de recente literatuur wordt een set aan criteria naar voren geschoven die in hun totaliteit van belang zijn voor kwaliteitsvolle evaluatie (Portael, 2013; Philips, Seghers, Versteden & Ysenbaert, 2013) die ook als doel heeft om effect te hebben op het leerproces:

- representativiteit
- reproduceerbaarheid
- authenticiteit
- cognitieve complexiteit
- geïntegreerdheid
- impact
- betrokkenheid van de leerling bij het evaluatieproces
- eerlijkheid
- transparantie
- verantwoording

Deze set aan criteria overstijgt de eerder binaire tegenstellingen die het toetsparadigma en assessmentparadigma met zich meebrengen door beide samen te brengen binnen dit geheel van criteria voor kwaliteitsvolle evaluatie. Dit geheel verwijst onder meer naar de klassieke psychometrische criteria via 'representativiteit' en 'reproduceerbaarheid'. Tegelijkertijd houdt deze set aan criteria ook rekening met de verschuiving die zich voordeed met betrekking tot wat voor doelen het onderwijs van vandaag wil nastreven (Baartman, Bastiaens, Kirschner, & Van der Vleuten, 2006). We zien dit terugkeren in de criteria 'authenticiteit' en 'cognitieve complexiteit' die aansluiten bij het steeds meer competentiegerichte denken in onderwijs. De overige criteria (impact, betrokkenheid, eerlijkheid, transparantie, verantwoording en geïntegreerdheid) achten we binnen de scope van dit onderzoek des te meer als relevant omdat ze rechtstreeks verband houden met de leerling zelf en hierdoor dus ook van belang zijn voor een evaluatiepraktijk die rekening houdt met de diversiteit onder leerlingen.

De **impact** van evaluatie verwijst naar de invloed die evaluatie heeft op het uiteindelijke leergedrag van leerlingen, waarbij een kwaliteitsvolle evaluatie een positief effect op het leergedrag van de leerling nastreeft (Segers & Tillema, 2011). De **betrokkenheid van de leerling** bij evaluatie verwijst

naar de mate waarin de leerling actief betrokken wordt bij het evaluatieproces. Een actieve rol van de leerling via bijvoorbeeld zelf- of peerevaluatie verhoogt de betrokkenheid en stimuleert de verantwoordelijkheid van de leerling over zijn/haar leerproces. Er zijn tal van aanwijzingen van positieve effecten van zelfevaluatie op het leerproces van de leerling (Black & Wiliam, 2009; Brookhart, 2007; Clark, 2012; Hattie & Timperley, 2007). Door het eigen leren of dat van *peers* te evalueren aan de hand van een specifieke set van criteria worden de zelfregulerende vaardigheden aangescherpt. Via zelf- en peerevaluatie worden leerlingen uitgedaagd om te reflecteren over hun eigen leerproces en leeruitkomsten waardoor ze meer eigenaar worden van hun leerproces.

Het criterium '**eerlijkheid**' verwijst naar het feit dat evaluatie billijk ofwel *fair* moet zijn voor alle leerlingen. Dit houdt in dat niet-relevante kenmerken van een leerling geen invloed uitoefenen op het uiteindelijke resultaat van de evaluatie. De evaluatie mag dus geen vertekening vertonen voor bepaalde groepen van leerlingen (Baartman et al., 2006). De **transparantie** verwijst naar de mate waarin de leerlingen geïnformeerd worden over de evaluatie. Meer specifiek gaat dit over de communicatie omtrent wat er geëvalueerd wordt, wanneer en op basis van welke criteria (Dierick & Dochy, 2001). De **verantwoording** verwijst naar de mate waarin de evaluatie gerechtvaardigd kan worden aan alle betrokkenen waaronder ook de leerlingen. Dit aspect onderstreept het belang van criteria en richtlijnen voor het toekennen van scores zodat er duidelijkheid is omtrent welke prestaties verwacht worden om bijvoorbeeld te slagen bij een bepaalde evaluatie (Van Petegem & Vanhoof, 2002). De **geïntegreerdheid**, ten slotte, verwijst naar de mate waarin de evaluatiepraktijk ingebed is in het ruimere onderwijsleerproces.

1.2. Opvattingen met betrekking tot evaluatie

Onderzoek toont aan dat de evaluatiecultuur op school er toe doet voor de evaluatiepraktijken die gehanteerd worden (Havnes, Smith, Dysthe, & Ludvigsen, 2012; Sach, 2015). In de literatuur wordt een onderscheid gemaakt tussen vier soorten opvattingen of concepties omtrent evaluatie. Een eerste verwijst naar de **bevordering van het onderwijsleerproces** en houdt in dat evaluatie voornamelijk beschouwd wordt als middel waarmee de leerkracht informatie verwerft over het leerproces van de leerlingen enerzijds en over de gebruikte onderwijsaanpak anderzijds (Black & Wiliam, 1998). Vandeyar and Killen (2007) toonden via hun onderzoek aan dat leerkrachten die evaluatie inzetten om zo feedback te kunnen geven aan leerlingen en tevens de eigen onderwijsaanpak bij te sturen, meer gebruik maken van evaluaties die geïntegreerd zijn in de dagelijkse onderwijspraktijk. Deze leerkrachten zullen formatieve evaluatie meer benadrukken dan summatieve en zullen leerlingen even goed belonen voor hun inzet als voor hun uiteindelijke resultaten.

Een tweede opvatting omtrent evaluatie beschouwt evaluatie als manier om mensen verantwoordelijk te houden voor het gestelde gedrag. Binnen deze '*accountability*-opvatting' wordt een onderscheid gemaakt tussen '**student-accountability**' enerzijds en '*schoolaccountability*' anderzijds. Het eerste stelt vooral de leerlingen verantwoordelijk voor hun leerproces en legt de nadruk op het behalen van bepaalde kwalificaties om bijvoorbeeld tot vervolgonderwijs te kunnen overgaan. In de praktijk zullen leerkrachten met dergelijke overtuiging omtrent evaluatie in hun praktijk sneller leerlingen vergelijken met medeleerlingen in plaats van hun individuele traject als referentie te nemen (Harris & Brown, 2009). Leerkrachten die hierop aansluiten, zullen in hun evaluatiepraktijk veel meer gebruik maken van formele evaluaties. De resultaten van laagpresterende leerlingen zullen eerder toegeschreven worden aan het gebrek aan capaciteiten van deze leerlingen of aan hun socio-economische

achtergrond dan aan de eigen onderwijspraktijk (Delandshere & Jones, 1999). De opvatting van **'school-accountability'** verwijst naar evaluatie als middel om de leerkracht, de school of, nog ruimer, het onderwijssysteem te beoordelen. Evaluaties geven – volgens deze benadering – de effectiviteit van leerkrachten, scholen of een onderwijssysteem weer. Tekorten bij leerlingen worden eerder verklaard vanuit de gebrekkige effectiviteit van leerkrachten of onderwijsinstanties.

Een laatste opvatting die Harris and Brown (2009) aanhalen is de **kritische** houding waarin (formele) evaluatie gezien wordt als irrelevant omdat ze noch voor de leerkracht, noch voor de leerling betekenisvol zijn. Formele evaluatiepraktijken worden als verwerpelijk beschouwd omdat er vooral negatieve aspecten aan verbonden worden. Deze kunnen variëren: evaluatie tast de autonomie van de leerkracht aan, het leidt de aandacht af van het eigenlijke leerproces van de leerling, het werkt contraproductief voor sommige leerlingen omdat het gepaard gaat met stress en faalangst, wat bijgevolg hun zelfvertrouwen kan schaden, de resultaten van evaluaties zijn niet altijd betrouwbaar, ... Leerkrachten die hiervan overtuigd zijn, zullen formatieve vormen van evaluatie vooral vermijden, en wat betreft de summatieve evaluaties zullen ze zich beperken tot dat wat verwacht of opgelegd wordt door externe actoren (Vandeyar & Killen, 2007).

2. Factoren op leerkrachtniveau

Aansluitend bij de multilevelbenadering willen we tevens onderzoeken welke factoren op micro- en mesoniveau bepalend zijn voor de evaluatiepraktijken van leerkrachten. In deze sectie zal ingegaan worden op het microniveau ofwel de factoren die verbonden zijn aan de leerkracht zelf: diens opvattingen, bekwaamheid, doelmatigheidsbeleving, ervaring, vooropleiding en professionalisering.

2.1. Opvattingen & attitudes leerkrachten

Het is algemeen bekend dat opvattingen van leerkrachten een invloed hebben op het dagdagelijks handelen in de klas. Ze zouden zelfs een grotere impact uitoefenen op de onderwijspraktijk dan de socio-economische context van de school (Griffiths, Gore, & Ladwig, 2006). Vandeyar and Killen (2007) toonden aan dat leerkrachten hun evaluatiepraktijk verschilt naargelang de opvattingen die ze erop na hielden. Leerkrachten die evaluatie beschouwden als manier om informatie te verkrijgen over het leerproces om zo de onderwijspraktijk bij te sturen, maakten in hun praktijk veel frequenter gebruik van formatieve evaluatievormen. Leerkrachten die evaluatie eerder zagen als een manier om leerlingen verantwoordelijk te houden voor hun leerproces, gebruikten vooral summatieve evaluaties. Hoewel verschillende auteurs (Nisbet & Warren, 2000) aantonen dat de evaluatiepraktijk van leerkrachten beïnvloed wordt door een overheersende visie op evaluatie, wijst onderzoek van Brown, Lake & Matters (2011) en Brown (2004) op de complexiteit van de relatie tussen opvattingen en evaluatiepraktijk. Brown (2004, 2011) toont aan dat leerkrachten meerdere – soms zelfs tegenstrijdige – visies met betrekking tot evaluatie kunnen hebben. Dit lijkt samen te hangen met het feit dat evaluatie op zich ook verschillende doelen dient, waardoor de leerkracht als evaluator ook verschillende rollen opneemt: ondersteuner van het leerproces, beoordelaar van het leren, motivator van het leren, rapporteur ten aanzien van verschillende actoren, etc. Deze meervoudige rollen maken de relatie tussen opvattingen en praktijken omtrent evaluatie complex. Bovendien wordt de relatie tussen opvattingen en evaluatiepraktijk ook doorkruist door de ruimere sociale en politieke context waarbinnen evaluatiepraktijken tot stand komen (Harris & Brown, 2009; Opre, 2015).

Bij uitbreiding kunnen de ruimere onderwijsopvattingen van leerkrachten samenhangen met hun opvattingen rond evaluatie en dus bepalend zijn voor de evaluatiepraktijk. Denessen (2000) maakt hiervoor een onderscheid tussen twee typen van opvattingen, ook wel ideologieën genoemd: leerstofgericht enerzijds en leerlinggericht anderzijds. Voor wat betreft de doelen binnen het onderwijs schuift de leerstofgerichte ideologie een patroon van opvattingen naar voren die de nadruk leggen op beroepsvoorbereiding. Inhoudelijk hecht deze ideologie ook veel belang aan kernvakken, zoals lezen, schrijven en rekenen. Onderwijs wordt in dit opzicht beschouwd als kennisoverdracht. Bovendien ligt de nadruk op de prestaties van de leerlingen. Er kan verwacht worden dat een leerstofgerichte ideologie zich in de evaluatiepraktijk vertaalt in een eerder productgerichte en summatieve benadering waarbij evaluatie eerder los staat van het onderwijsleerproces (cf. gescheiden model). De leerlinggerichte ideologie daarentegen, benadrukt veel meer het doel van individuele ontplooiing en maatschappelijke voorbereiding. In deze benadering van onderwijs krijgen creatieve en sociale vakken meer aandacht en ligt de nadruk veel meer op het leerproces zelf en in mindere mate op de uitkomsten van dit proces. Deze benadering van onderwijs sluit veel meer aan bij een brede visie op evalueren waarin een evenwicht nagestreefd wordt tussen proces- en productevaluatie. Bovendien beperkt het evalueren zich niet tot kernvakken, maar staat evaluatie meer in het teken van de totale ontwikkeling van de leerling.

Met het oog op het nagaan in welke mate de evaluatiepraktijk van leerkrachten tegemoet komt aan de diversiteit onder leerlingen, willen we zicht krijgen op de houding van leerkrachten ten aanzien van diversiteit. Onderzoek omtrent evaluatiepraktijken in de context van de apartheid in Zuid-Afrika gaf aan dat leerkrachten in hun evaluatiepraktijk niet geneigd waren om aanpassingen te doen ten aanzien van leerlingen die op vlak van taal of cultuur niet beantwoordden aan de heersende schoolcultuur (Vandeyar & Killen, 2007). Integendeel, de evaluatiepraktijken leken veeleer gericht op het bestendigen van de verschillen tussen leerlingen door het opdragen van de schoolcultuur als de te bereiken standaard. Deze studie liet tevens zien dat deze evaluatiepraktijken samen hingen met opvattingen omtrent evaluatie die eerder steunen op het verantwoordingsgerichte perspectief op evaluatie in plaats van het ontwikkelingsgerichte perspectief. Een mogelijke verklaring kan gevonden worden in de mate waarin leerkrachten overtuigd zijn van de maatschappelijke verantwoordelijkheid van onderwijs om rekening te houden met diversiteit. Leerkrachten die hiervan overtuigd zijn, zullen meer geneigd zijn om in hun eigen klaspraktijk hiermee rekening te houden (Lee & Loeb, 2000). Er zullen in dat geval ook meer inspanningen geleverd worden om alle leerlingen kwaliteitsvol onderwijs te bieden (Fischman, DiBara, & Gardner, 2006) en dus ook de evaluatie zo in te richten dat die de leer- en ontwikkelingskansen van leerlingen maximaliseert.

2.2. Bekwaamheid & doelmatigheidsbeleving

Naast het geheel aan opvattingen en attitudes, speelt de bekwaamheid van leerkrachten een rol voor de effectiviteit van hun handelen. Leerkrachten die het gevoel hebben dat ze een invloed hebben op het leren en het gedrag van leerlingen zullen beter in staat zijn om te variëren in evaluatiestrategieën en de evaluatie af te stemmen op het ruimere onderwijsleerproces.

De bekwaamheid van leerkrachten om aan differentiatie te doen en om de gepaste evaluatie te ontwikkelen of te selecteren kan gerelateerd zijn aan de gevolgde vooropleiding en/of de mate waarin men inzet op bijkomende professionaliseringsinitiatieven. Leerkrachten worden over het algemeen

beperkt voorbereid op het evalueren van leerlingen (Sluijsmans & Struyven, 2014). Zo concludeert het TALIS-rapport (Teaching and Learning International Survey), waarbij 23 OESO-landen vergeleken werden, dat 15,7% van de leerkrachten in het lager secundair onderwijs een sterke behoefte voelt om zich te professionaliseren rond leerlingenevaluatie. Vlaanderen situeert zich in deze ranking van 23 landen op plaats 9, net onder het gemiddelde (OECD, 2009, 2013). Gelijkaardige bevindingen werden vastgesteld bij het jaarlijks rapport van de inspectie waarbij beginnende leerkrachten gevraagd werd om aan te geven welke leerkrachtcompetentie ze als het meest moeilijk ervaren in de praktijk. Hieruit bleek dat het ontwikkelen en selecteren van gepaste evaluatie in de top drie van de als meest moeilijk ervaren vaardigheden terecht kwam. Nog verontrustender in hetzelfde rapport is de vaststelling dat directies leerlingenevaluatie als één van de minder belangrijke competenties zien voor beginnende leerkrachten (Onderwijsinspectie, 2006). Dit houdt een risico in voor de ondersteuning en verdere professionalisering van beginnende leerkrachten op vlak van leerlingenevaluatie.

Deze bevindingen werden bevestigd in het kwalitatieve luik van het huidige onderzoek waarin vastgesteld werd dat heel wat leerkrachten zich onzeker voelen met betrekking tot hun evaluatiepraktijk. Bovendien blijkt die onzekerheid toe te nemen als het gaat om het rekening houden met verschillen tussen leerlingen op momenten van evaluatie (Ysenbaert et al., 2018). Dit werd in het OESO-rapport ook als grootste uitdaging geformuleerd: *“The key challenge is to ensure that rich assessment opportunities are systematically offered to all students regardless of characteristics like behavior, gender, special educational needs, immigrant background, first language, overall academic achievement or verbal ability”* (OECD, 2013).

2.3. Ervaring

Nye, Konstantopoulos, and Hedges (2004) wezen er al op dat de rol van de leerkracht er toe doet aangezien leerkrachten verschillen in de mate waarin ze leerwinst bereiken bij hun leerlingen. Ervaring van leerkrachten zou mogelijks één van de factoren kunnen zijn om het verschil in effectiviteit tussen leerkrachten te verklaren. Sommige studies bevestigen dat de ervaring van de leerkracht een positief effect heeft op de leeruitkomsten van leerlingen (Clotfelter, Ladd, & Vigdor, 2006). Hattie (2003) echter, stelt dat er een belangrijk verschil is tussen enerzijds ervaren leerkrachten en anderzijds wat Hattie ‘experten’ noemt, namelijk leerkrachten die bekwaam zijn in hun vak, ongeacht het aantal jaren ervaring dat ze hebben. Het aantal jaar ervaring van een leerkracht zou dus niet noodzakelijk samenhangen met zijn/haar expertise in het leraar zijn.

Specifiek met betrekking tot ‘evaluatie’ zien we in de casestudies dat leerkrachten er op wezen dat het aantal jaar ervaring hen doorgaans beter in staat stelt om evaluatie en instructie sterker op elkaar af te stemmen. Leerkrachten die al jaren les gaven, vonden dat hun evaluatiepraktijk in het begin van hun loopbaan bijvoorbeeld nog sterk geïsoleerd stond van de algemene lespraktijk.

3. Factoren op schoolniveau

Onderzoek naar de evaluatiepraktijk van leerkrachten toont aan dat deze niet uitsluitend beïnvloed worden door factoren die verbonden zijn aan de leerkrachten (microniveau), maar eveneens aan factoren die zich situeren op meso- of schoolniveau. Zo is het beleidsvoerend vermogen van de school een cruciale factor in hoe de leerkrachten (ondersteund worden om) hun evaluatiepraktijk vorm (te) geven, daarnaast kunnen de kenmerken van de school, zoals de leerlingenpopulatie en het

studieaanbod, een impact hebben op de schoolcultuur omtrent evaluatie en dus ook op de evaluatiepraktijken van leerkrachten.

3.1. Beleidsvoerend vermogen

Naast het feit dat onderzoek al aantoonde dat de leerkracht er toe doet, bestaat er evenveel evidentie voor het feit dat de school er toe doet. Dit vloeit voort uit de vaststelling dat scholen die over dezelfde mate van beleidsruimte beschikken, deze beleidsruimte toch verschillend lijken aan te wenden. Zelfs al streven scholen ernaar om dezelfde doelen te bereiken, toch blijken ze te verschillen in de manier waarop ze deze doelen nastreven. Met betrekking tot de pedagogisch-didactische invulling is het bovendien zo dat de scholen in Vlaanderen over een grote mate van autonomie beschikken. Dit betekent dat ze vrij zijn in hoe het evaluatiebeleid en de evaluatiepraktijken vorm krijgen. Dit vergt echter een zekere mate van beleidsvoerend vermogen om die beleidsruimte zo in te vullen opdat de kwaliteit van het onderwijs bevorderd wordt. Van den Berg (2001) merkt op dat indien veel beleidsruimte gecombineerd wordt met een gebrekkig beleidsvoerend vermogen, scholen sneller vervallen in passiviteit en onderbenutting van deze beleidsruimte waardoor hun werking gekenmerkt wordt door conservatisme (Van den Berg, 2001).

Met betrekking tot het beleidsvoerend vermogen van scholen onderscheiden (Vanhoof, Deneire, & Van Petegem, 2011) een achttal indicatoren:

- doeltreffende communicatie
- samenwerking en ondersteunende relaties
- betrokkenheid via gedeeld leiderschap
- gezamenlijk doelgerichtheid
- responsief ten aanzien van zowel interne als externe verwachtingen
- innoverend vermogen
- integratie van beleidsinitiatieven
- reflectief vermogen

Uit het voorafgaand casestudieonderzoek naar evaluatiebeleid en – praktijk van scholen, bleken voornamelijk de indicatoren doelgerichtheid, doeltreffende communicatie en reflectief vermogen op de voorgrond te treden bij die scholen die een sterker uitgewerkt evaluatiebeleid hebben.

Elke school heeft een visie op onderwijs en leren en daarmee samenhangend een te realiseren doel. Aangezien onderwijs pas effectief en succesvol is indien de evaluatiepraktijk ook in lijn is met de onderwijspraktijk, dient de visie op evaluatie voort te vloeien uit de ruimere visie op onderwijs en leren (James, 2006). Het tot stand komen van een **gezamenlijke doelgerichtheid** over het evalueren van leerlingen vergt van schoolteams dat ze als team het hoofd buigen over een visie rond leerlingenevaluatie en over welke doelen men hierbij voorop wil stellen. Dit vraagt dat men de verschillende visies hieromtrent tot overeenstemming brengt. De gezamenlijke doelgerichtheid wordt gekenmerkt door overeenstemming die zich uit in de wijze waarop leerlingen geëvalueerd worden. Om tot gezamenlijke doelgerichtheid inzake leerlingenevaluatie te komen moeten de leerkrachten betrokken worden (Troudi, Coombe, & Al-Hamliy, 2009). Bovendien speelt de schoolleider een cruciale rol in het installeren van evaluatiecultuur waar eenzelfde visie, doelen en verwachtingen gedeeld worden (Havnes et al., 2012; Sach, 2015).

De betrokkenheid van de leerkrachten kan gefaciliteerd worden door een **doeltreffende communicatie** op te zetten over leerlingenevaluatie. Een doeltreffende communicatie houdt in dat de wederzijdse beïnvloeding van leden binnen het team leidt tot kennisconstructie en bijgevolg ook tot doelgericht handelen in ieders praktijk (Vanhoof et al., 2011). De communicatie dient op twee niveaus te gebeuren met enerzijds een verticale informatiedoorstroom die verwijst naar de wisselwerking tussen de schoolleiding en het team van leerkrachten. Anderzijds dient er gewaakt te worden over de communicatie op horizontaal niveau tussen leerkrachten onderling. Uit de casestudies bleek dat communicatie een cruciale factor bleek te zijn als het gaat over de evaluatie van leerlingen, en des te meer indien het bleek te gaan om leerlingen met specifieke noden (Ysenbaert et al., 2018).

Het **reflectief vermogen** van een school blijkt bij te dragen tot het beleidsvoerend vermogen. Het reflectief vermogen fungeert als motor voor de kwaliteitsbewaking van een school: via reflectie kan een school bijleren over de gangbare evaluatiepraktijken en die eventueel in vraag stellen. In de casestudies werd namelijk vastgesteld dat evaluatiepraktijken in veel scholen berusten op afspraken die in het verleden gemaakt werden, maar in tussentijd nog maar weinig verankerd zijn in de hedendaagse verwachtingen van kwaliteitsvol onderwijs. Zo stouwen zaken als de frequentie waarmee er geëvalueerd wordt en de manier waarop er gerapporteerd wordt nog vaak op gewoonte. Reflectieve scholen durven de eigen evaluatiepraktijken kritisch bekijken. Vastgeroeste gewoontes en routines, in het algemeen en ook op vlak van evalueren zijn nefast voor het reflectief vermogen. Wanneer een school erin slaagt om te reflecteren en zo de sterke en zwakke punten in de evaluatiepraktijken kan identificeren, draagt dit bij tot het beleidsvoerend vermogen (Portael, 2013). ‘Evaluatie’ kan hier op het schoolniveau ook een doel hebben: namelijk het bijsturen van de onderwijs- én evaluatiepraktijk. In de literatuur wordt het gebruik van leerlingresultaten om de onderwijspraktijk bij te sturen in verband gebracht met beter presterende scholen (Robinson, Lloyd, & Rowe, 2008). Dit werd bevestigd in de cases die onderzocht werden en waaruit bleek dat scholen die evaluatie aanwenden in het kader van interne kwaliteitsbewaking meer geneigd zijn om de vooruitgang van hun leerlingen te monitoren. Deze monitoring moedigt scholen bijvoorbeeld aan om hun onderwijsorganisatie en professionaliseringsbeleid te herdenken in functie van het maximaliseren van de leeruitkomsten van hun leerlingenpubliek.

3.2. Studieaanbod

Volgens het onderzoek van Remesal (2007, 2011) en Brown et al. (2011) verschillen de opvattingen omtrent evaluatie bij leerkrachten naargelang het onderwijsniveau waarin ze les geven. Leerkrachten in het secundair onderwijs leggen meer de nadruk op het verantwoordingsaspect, terwijl leerkrachten in het lager onderwijs evaluatie veel meer zien als een middel om het onderwijsleerproces bij te sturen. De resultaten uit het voorgaande casestudie-onderzoek lijken dit te bevestigen (Ysenbaert et al., 2018). Deze verschillen in onderwijsniveau worden eerder verklaard door de verschillende verwachtingen in beide onderwijsniveaus, dan door het onderwijsniveau zelf.

Onderzoek naar evaluatiepraktijken van leerkrachten toonde al aan dat dit kan verschillen naargelang het vak of het domein waar dit betrekking op heeft: leerkrachten zouden voor het vak wiskunde, bijvoorbeeld, veel meer gebruik maken van meerkeuzetoetsen en reproductievragen dan dat leerkrachten voor een taalvak dit doen (Duncan et al., 2009; Duncan & Noonan, 2007). Op deze manier kunnen er eventueel ook verschillen waargenomen worden in evaluatiepraktijken tussen de verschillende onderwijsvormen omdat de aard van de vakken in bijvoorbeeld een aso-richting verschilt van deze in een bso-richting.

Het casestudieonderzoek wees op de mogelijke invloed van de manier waarop een school haar onderwijs organiseert voor de evaluatiepraktijk van leerkrachten (Ysenbaert et al., 2018). Zo blijken scholen die afwijken van het jaarklassensysteem beter in staat te zijn om op momenten van evaluatie te differentiëren in doelen en om leerlingen veel meer te betrekken bij evaluatie via zelf- en peerevaluatie. Hierbij wordt opgemerkt dat de manier waarop het onderwijs georganiseerd wordt eerder een gevolg is van visie die men erop na houdt met betrekking tot onderwijs en leren, waardoor de evaluatiepraktijken tevens verschillend ingevuld worden.

3.3. Leerlingencompositie

Onderzoek toonde reeds aan dat de verwachtingen die leerkrachten hebben ten aanzien van hun leerlingen cruciaal zijn om leerlingen te stimuleren in hun leerproces (Rosenthal & Jacobson, 1968). Lage verwachtingen van leerkrachten blijken zich ook te vertalen in de wijze waarop ze leerlingen evalueren. McMillan, Myran, and Workman (2002) ontdekten dat leerkrachten hun evaluatiepraktijken aanpassen naargelang hun leerlingenpubliek. Zo zouden sterk presterende leerlingen vaker evaluaties krijgen die de hogere orde denkvaardigheden aanspreken. Laag-presterende leerlingen zouden hier minder toe uitgedaagd worden omdat voor deze groep eerder evaluatievormen gehanteerd worden die peilen naar kennisreproductie. Stevens (2007) toonde aan dat de evaluatiepraktijken van leerkrachten beïnvloed worden door druk van buitenaf en verwachtingen vanuit de bredere omgeving waardoor ze in hun evaluatie bijvoorbeeld lagere verwachtingen gaan stellen voor specifieke groepen van leerlingen. Dit kan het geval zijn in functie van het behouden van het leerlingenaantal om een bepaalde studierichting te kunnen blijven behouden.

Hoofdstuk III Methodologie

1. Onderzoeksopzet & onderzoeksvragen

Dit survey-onderzoek beoogt een dubbel doel: enerzijds willen we meer inzicht verwerven in hoe de evaluatiepraktijken van leerkrachten eruit zien. Anderzijds willen we nagaan hoe deze evaluatiepraktijken tot stand komen.

Met betrekking tot het eerste doel streven we ernaar om de eerder ontwikkelde modellen voor evaluatiepraktijk (Ysenbaert et al., 2017, 2018) verder te verfijnen. Meer specifiek willen we nagaan in welke mate de drie modellen voor evaluatiepraktijk samenhangen met de kwaliteitscriteria 'eerlijkheid', 'transparantie', 'betrokkenheid', 'impact' en 'verantwoording'. Aangezien deze criteria zich meer focussen op de leerling zelf verwachten we dat evaluatiepraktijken die hierop inzetten positief zullen samenhangen met het geïntegreerde model voor evaluatiepraktijk. Bovendien willen we ook nagaan hoe het rekening houden met deze kwaliteitsaspecten samenhangt met de twee overige modellen (het gescheiden en het semi-geïntegreerd model). Door de verhouding te onderzoeken tussen de modellen voor evaluatiepraktijken enerzijds en de aandacht voor de verschillende evaluatiecriteria anderzijds streven we een beter begrip na van deze modellen. Tegelijkertijd laat dit ons toe om na te gaan in welke mate deze modellen al dan niet te situeren zijn op een continuüm.

Naast de aandacht voor de verschillende kwaliteitsaspecten willen we ten slotte nagaan of een geïntegreerde evaluatiepraktijk wel degelijk meer rekening houdt met de diversiteit onder leerlingen. We gaan dit na door de samenhang te onderzoeken tussen de modellen voor evaluatiepraktijk en de wijze waarop leerkrachten in hun algemene onderwijspraktijk inzetten op differentiatie. In de casestudies die dit survey-onderzoek vooraf gingen, waren er al voorzichtige aanwijzingen voor het verband tussen gedifferentieerde klaspraktijk en gedifferentieerde evaluatiepraktijk. In deze cases waar klaspraktijken gekenmerkt werden door sterkere vormen van binnenklasdifferentiatie, werd er ook veel meer ingezet op het breed evalueren van leerlingen (Ysenbaert et al., 2018).

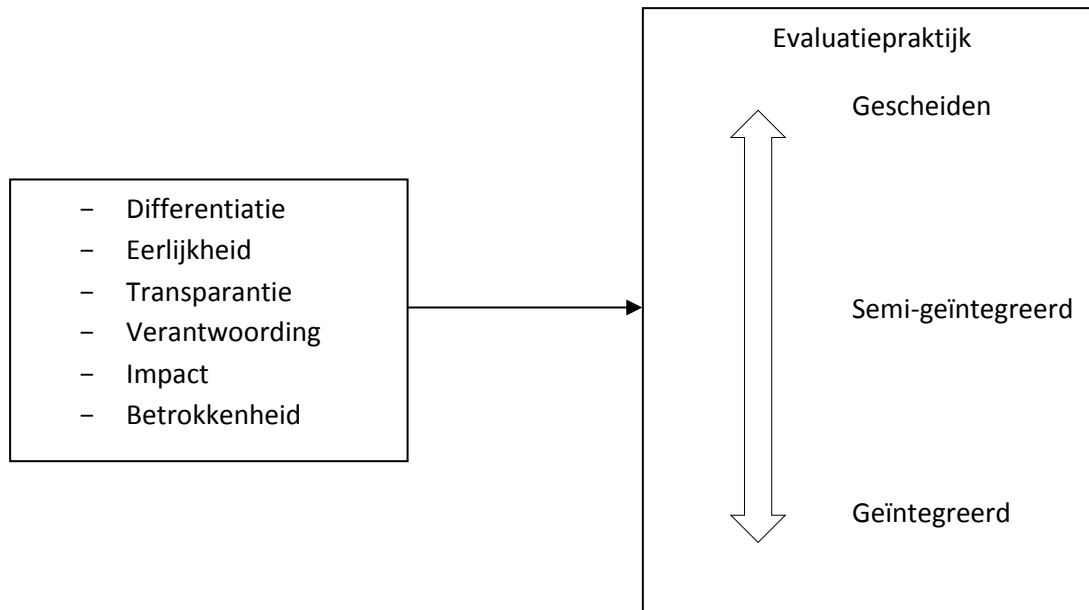


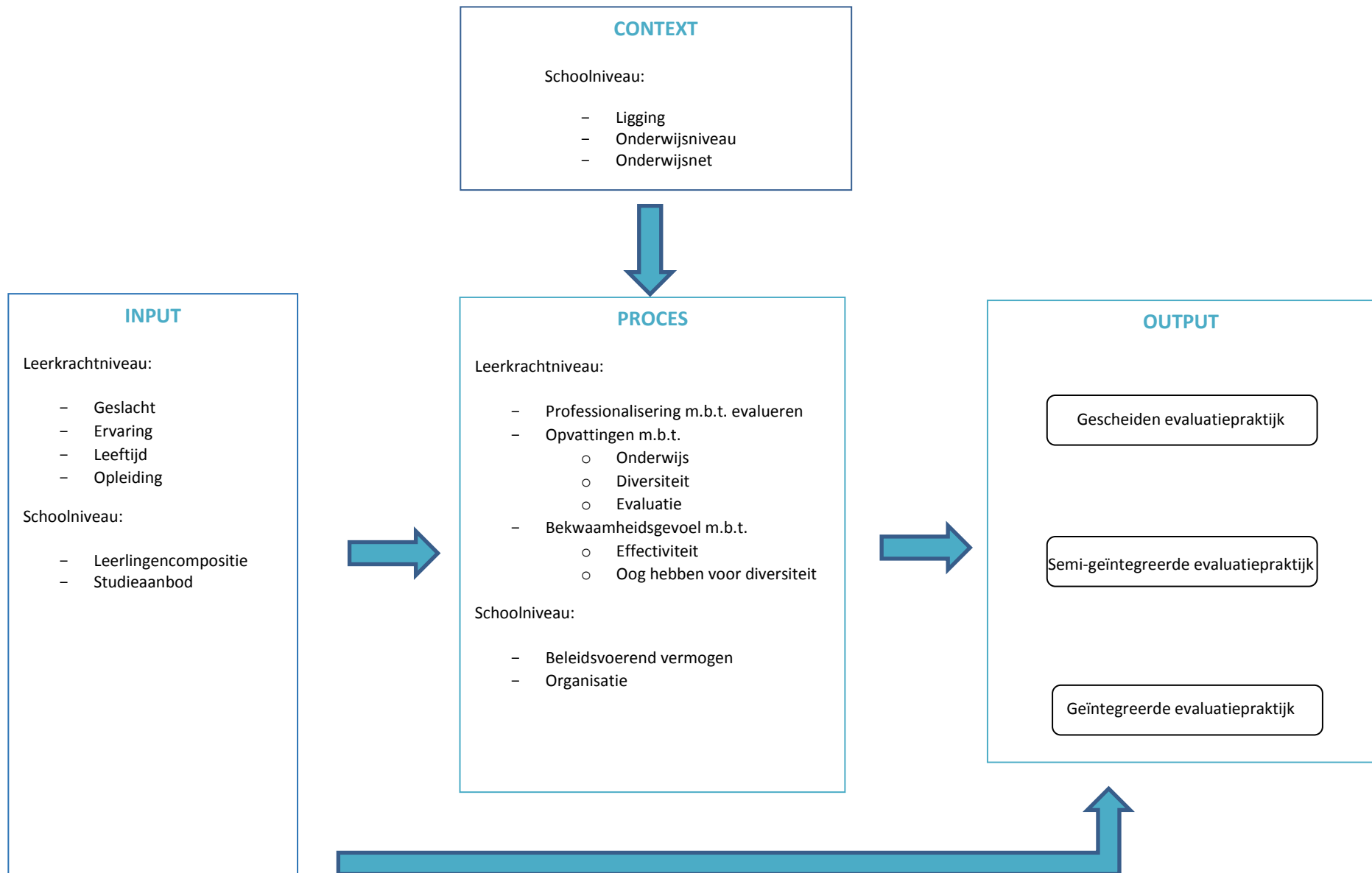
Fig. 5: Conceptueel model voor evaluatiepraktijk

Aan de hand van bovenstaand conceptueel model voor evaluatiepraktijk willen we de drie modellen voor evaluatiepraktijk - die ontwikkeld werden op basis van de literatuurstudie en reeds gehanteerd werden bij het casestudie-onderzoek - verder gaan verfijnen. Concreet willen we via het survey-onderzoek volgende onderzoeksvragen afoetsen:

- Hangt een geïntegreerde evaluatiepraktijk positief samen met een sterkere differentiatiepraktijk in het algemeen en met de verschillende kwaliteitsaspecten voor evaluatie zoals 'eerlijkheid', 'transparantie', 'verantwoording', 'impact' en 'betrokkenheid'?
- Hangt een gescheiden evaluatiepraktijk negatief samen met een sterkere differentiatiepraktijk in het algemeen en met de verschillende kwaliteitsaspecten voor evaluatie zoals 'eerlijkheid', 'transparantie', 'verantwoording', 'impact' en 'betrokkenheid'?
- Kunnen de drie modellen voor evaluatiepraktijk geplaatst worden op een continuüm waarbij de effecten voor de semi-geïntegreerde evaluatiepraktijk minder uitgesproken zijn dan bij de gescheiden of geïntegreerde evaluatiepraktijk?
- Is er sprake van een schoolcultuur met betrekking tot de wijze waarop er geëvalueerd wordt? Zijn de evaluatiepraktijken van leerkrachten op eenzelfde school dermate gedeeld dat we kunnen stellen dat de evaluatiepraktijk op een school gekenmerkt wordt door één van de drie modellen?

Met betrekking tot het tweede doel van dit survey-onderzoek willen we ook inzicht krijgen in hoe deze praktijken tot stand komen. Meer specifiek willen we nagaan welke de bepalende factoren zijn op leerkracht- en schoolniveau voor een gescheiden, semi-geïntegreerde en geïntegreerde evaluatiepraktijk. Hiertoe hanteren we het CIPO-model waarin het hierboven omschreven theoretisch kader in ondergebracht wordt. Dit model is opgebouwd volgens de indeling *Context-Input-Proces-Output*. Het CIPO-model wordt binnen onderwijseffectiviteitsonderzoek vaak naar voren geschoven om onderzoeksvragen te concipiëren en te ordenen (Scheerens, 1990). Binnen het huidige onderzoek

waarin we op zoek zijn naar bepalende factoren voor de evaluatiepraktijken in scholen, wordt dit als *output* beschouwd. Via het survey-onderzoek willen we nagaan hoe de evaluatiepraktijken verklaard kunnen worden door mogelijke *context-, input- en proceskenmerken*:



Concreet willen we aan de hand van bovenstaand model een antwoord zoeken op volgende onderzoeksvragen:

- Welke opvattingen met betrekking tot evaluatie kunnen we onderscheiden bij de bevroegde leerkrachten?
- Zijn de opvattingen van de leerkrachten op eenzelfde school dermate gedeeld dat we kunnen spreken van een schoolcultuur op vlak van evaluatie?
- Hoe hangen de opvattingen van leerkrachten samen met de evaluatiepraktijk van leerkrachten? We verwachten dat een sterke overtuiging van de bevordering voor het onderwijsleerproces positiever zal samenhangen met het (semi-)geïntegreerde model voor evaluatiepraktijk. Bij het gescheiden model verwachten we eerder een negatief verband.
- In welke mate zijn de opvattingen van leerkrachten met betrekking tot:
 - o Onderwijs (leerstofgericht & leerlinggericht)
 - o Evaluatie
 - o Diversiteitbepalend voor de evaluatiepraktijk van leerkrachten?
- In welke mate is de bekwaamheid van leerkrachten
 - o Self-efficacy
 - o Oog hebben voor diversiteit
 - o Mate waarin men inzet op professionalisering
 - o Gevolgde vooropleiding
 - o Ervaringbepalend voor de evaluatiepraktijk van leerkrachten?
- In welke mate is het beleidsvoerend vermogen bepalend voor de evaluatiepraktijk van leerkrachten?
- In welke mate is het studieaanbod van de school bepalend voor de evaluatiepraktijk van leerkrachten?
- In welke mate is de leerlingencompositie van de school bepalend voor de evaluatiepraktijk van leerkrachten.

2. Steekproef

2.1. Basisonderwijs

2.1.1. Steekproeftrekking

Voor het basisonderwijs werd geopteerd om met de 30 scholen verder te werken die in 2017 betrokken waren bij de dataverzameling van onderzoekslijn 1.5. Bij de contactname van de scholen in 2017 werd meegedeeld dat het voor de school om een tweeledig onderzoek zou gaan, een eerste luik omtrent taalstimulering en een tweede luik in 2019 omtrent evaluatiebeleid en –praktijk.

2.1.2. Contactname

In de periode januari-april 2019 werden de 30 basisscholen gecontacteerd. In eerste instantie ontvingen de scholen een mail met bijhorend een antwoordformulier waarop men kon aangeven wie in de school als contactpersoon aangesteld werd voor de verdere opvolging van de afname van de survey. Bij 8 van de 30 scholen werd er positief gereageerd op dit e-mailbericht. De overige scholen kregen een herinneringsmail, dit gecombineerd met een, vaak herhaaldelijke (tot maximaal negen keer) telefonische contactname. Dit resulteerde in de bevestiging van nog 21 scholen. Drie scholen lieten weten dat ze niet meer wensen deel te nemen omdat ze reeds bij andere projecten/onderzoeken betrokken zijn.

2.1.3. Definitieve steekproef

Voor de operationalisering van ‘leerlingencompositie’ hanteren we het percentage leerlingen van wie de moeder laagopgeleid (geen diploma hoger secundair onderwijs) is. De definitieve steekproef bestaat uit 12 scholen die gekenmerkt worden door een eerder homogeen samengesteld leerlingenpubliek, waarbij de meerderheid van de leerlingen een hoogopgeleide moeder heeft (hoogstens 20% van de leerlingen heeft een laag opgeleide moeder). Zes scholen zijn qua leerlingencompositie eveneens homogeen samengesteld, maar de meerderheid (> 60%) van de leerlingen heeft een laagopgeleide moeder. Negen scholen zijn qua leerlingencompositie heterogeen samengesteld: 30 tot 50 % van de leerlingen heeft een laagopgeleide moeder.

Tabel 3.1

Leerlingencompositie in de definitieve steekproef basisonderwijs

Leerlingencompositie	N
Homogeen hoge SES	12
Heterogeen	9
Homogeen lage SES	6

Ongeveer 3/4 van de scholen behoort tot het *vrij gesubsidieerd onderwijs*. Vijf scholen behoren tot het *gemeenschapsonderwijs* en twee scholen tot het *officieel gesubsidieerd onderwijs*.

Tabel 3.2

Spreiding onderwijsnetten in de definitieve steekproef basisonderwijs

Net	N
GO	5
OGO	2
VGO	20

De spreiding over de provincies is evenwichtig verdeeld, met uitzondering van de provincie Limburg die iets sterker vertegenwoordigd is.

Tabel 3.3

Spreiding provincies in de definitieve steekproef basisonderwijs

Provincie	N
Antwerpen	5
Oost-Vlaanderen	6
West-Vlaanderen	4
Limburg	7
Vlaams-Brabant	5

2.2. Secundair onderwijs

2.2.1. Steekproeftrekking

Voor de scholen van het secundair onderwijs werd, in tegenstelling tot het basisonderwijs, een nieuwe gestratificeerde toevalssteekproef getrokken van 51 secundaire scholen (vestigingsplaatsen). Hiervoor werd gebruik gemaakt van het populatiedatabestand van S. Groenez dat ter beschikking gesteld werd via de SONO-sharepoint.

Bij de steekproeftrekking werd in eerste instantie gestreefd naar voldoende variatie in leerlingencompositie. Hiervoor werd gekeken naar de cijfers voor SES die aangeven welk percentage van leerlingen een moeder heeft die laagopgeleid (geen diploma hoger secundair onderwijs) is of een schooltoelage ontvangt. Zo werden er 17 scholen geselecteerd waarvan de leerlingenpopulatie getypeerd wordt door een 'homogeen lage SES'. Het gaat hierbij om scholen waarvan hoogstens 26% van de leerlingen een hoogopgeleide moeder heeft of geen schooltoelage ontvangt. Daarnaast werden er 17 scholen geselecteerd waarvan minstens 79% van de leerlingen een hoogopgeleide moeder heeft of geen schooltoelage ontvangt. Een derde groep van 17 scholen werd geselecteerd op basis van de heterogene leerlingencompositie, ofwel scholen waarvan minstens 50% en hoogstens 57% van de leerlingen een laagopgeleide moeder heeft of een schooltoelage ontvangt.

In tweede instantie werden scholen geselecteerd op basis van onderwijsnet, studieaanbod en provincie.

Voor elke school die niet wenste deel te nemen, werd telkens een vervangschool gezocht. Hierbij werd erop toegezien dat de vervangscholen in eerste instantie beantwoorden aan het selectie criterium van leerlingencompositie. Indien bleek dat een bepaald(e) net, provincie of studieaanbod over- of ondervertegenwoordigd zou zijn, dan werd hiermee rekening gehouden bij het selecteren van vervangscholen.

2.2.2. Contactname

In de periode van januari tot aan de paasvakantie begin april 2019 werden er in totaal 95 secundaire scholen gecontacteerd. In eerste instantie ontvingen de 51 scholen uit de initiële steekproef een uitnodiging via mail met bijhorend een antwoordformulier waarop men kon aangeven wie in de school als contactpersoon aangesteld wordt indien men wenst deel te nemen aan het onderzoek. Indien men niet wenst in te tekenen voor deelname werd gevraagd naar de reden waarom men hier niet op wou ingaan. De communicatie via mail bij deze 51 scholen leverde twee bereidwillige scholen op. De overige scholen werden telefonisch gecontacteerd. Voor de scholen die niet wensten deel te nemen werd telkens een vervangschool gezocht, waardoor er uiteindelijk 95 scholen aangeschreven werden.

Het eerste contact met de vervangscholen verliep eveneens via mail om vervolgens de contactname verder te zetten via telefoon (van nul tot negen keer per school), gecombineerd met het versturen van herinneringsmails. In totaal werden voor het overtuigen van scholen met een aanbod van gewoon secundair onderwijs 156 mails verstuurd en 320 telefoontjes gepleegd. Gemiddeld werd elke school vijf keer gecontacteerd (via mail/telefoon) met de vraag of men wou deelnemen. Begin april (net vóór de paasvakantie) werd de fase van contactname beëindigd met een totaal van 30 scholen die bereid gevonden werden om deel te nemen aan het onderzoek en 65 scholen die antwoordden dat ze niet wensten deel te nemen.

De meest voorkomende reden die een school gaf om niet deel te nemen aan het onderzoek was overbevraging voor onderzoeken en projecten of omdat ze reeds deelnamen aan een ander onderzoek of project (N = 57). Enkele directies gaven hierbij aan dat ze daarom volgende 'strategieën' hanteren: uitsluitend ingaan op vragen van oud-leerlingen, enkel op deze vragen die aansluiten bij beleidsprioriteiten, bij aanvang van het schooljaar intekenen op enkele onderzoeken/projecten en gedurende het schooljaar geen nieuwe bijkomende vragen meer opnemen. Enkele scholen zien af van deelname omwille van het feit dat ze zich in een veranderingstraject bevinden, al dan niet door de hervorming voor volgend schooljaar (N = 10). Twee scholen gaven aan dat ze niet deelnemen omwille van een doorlichting (N = 2). Eén school verwees naar het feit dat de algemene werklast van de leerkrachten al te hoog lag (N = 1). Eén school, ten slotte, wou heel specifiek inzetten op evaluatiebeleid en -praktijk en had hierrond net een bevraging gedaan bij het team, waardoor men het team niet nog eens wou belasten met een bevraging rond hetzelfde thema.

De scholen die niet wensen deel te nemen variëren onderling naar leerlingencompositie, onderwijsnet, provincie en studieaanbod. In vergelijking met de definitieve steekproef met deelnemende scholen zien we dat er vooral uitval is bij die scholen die gekenmerkt worden door een hoge SES-leerlingencompositie en een studieaanbod met aso of eerste graad.

Tabel 3.4

Leerlingencompositie niet-deelnemende scholen secundair onderwijs

Compositie	N
Homogeen lage SES	20
Heterogeen	19
Homogeen hoge SES	26

Tabel 3.5

Spreading onderwijsnetten niet-deelnemende scholen secundair onderwijs

Net	N
GO	18
OGO	3
VGO	44

Tabel 3.6

Spreading provincie niet-deelnemende scholen secundair onderwijs

Provincie	N
Antwerpen	22
Oost-Vlaanderen	16
West-Vlaanderen	10
Limburg	7
Vlaams-Brabant	10

Tabel 3.7

Studieaanbod niet-deelnemende scholen secundair onderwijs

Studieaanbod	N
aso	15
aso-tso	3
aso-tso-bso	13
kso-tso-bso	2
tso-bso	13
1e graad	19

2.2.3. Definitieve steekproef

De definitieve steekproef in het secundair onderwijs omvat in totaal 30 scholen.

Met betrekking tot de diversiteit onder de leerlingen zien we dat de definitieve steekproef 12 scholen bevat waarvan de leerlingenpopulatie getypeerd wordt door een 'homogeen lage SES'. Het gaat hierbij om scholen waarvan hoogstens 26% van de leerlingen een hoogopgeleide moeder heeft of geen schooltoelage ontvangt. Daarnaast zijn er zes scholen waarvan minstens 79% van het leerlingenpubliek een hoogopgeleide moeder heeft of geen schooltoelage ontvangt. Een derde groep van 12 scholen wordt gekenmerkt door een heterogene leerlingensamenstelling, waarbij minstens 50% en hoogstens 57% van de leerlingen een laagopgeleide moeder heeft of een schooltoelage ontvangt.

Tabel 3.8

Leerlingensamenstelling in de definitieve steekproef secundair onderwijs

Leerlingensamenstelling	N
Homogeen lage SES	12
Heterogeen	12
Homogeen hoge SES	6

Wat betreft het onderwijsnet gaat het vooral om scholen uit het *vrij gesubsidieerd onderwijs* (N=20). Er zijn acht scholen uit het *gemeenschapsonderwijs* en twee scholen uit het *officieel gesubsidieerd onderwijs*.

Tabel 3.9

Spreiding onderwijsnetten in de definitieve steekproef secundair onderwijs

Net	N
GO	8
OGO	2
VGO	20

Het grootste aandeel van de scholen heeft een studieaanbod in tso-bso (N=11), zeven scholen hebben een aanbod in zowel, aso, tso als bso. Vier scholen hebben een studieaanbod in de onderwijsvormen aso-tso. Drie scholen hebben enkel een aanbod in de eerste graad waarbij er geen onderscheid is in onderwijsvormen. Twee scholen bieden onderwijs in de aso-kso en twee scholen in kso-tso-bso. Eén school, ten slotte, biedt uitsluitend onderwijs aan in het tso.

Tabel 3.10
Studieaanbod in de definitieve steekproef secundair onderwijs

Studieaanbod	N
aso-tso-bso	7
aso-tso	4
aso-kso	2
tso	1
tso-bso	11
kso-tso-bso	2
1e graad	3

De verdeling van de scholen over de provincies heen laat zien dat de provincie Limburg niet vertegenwoordigd is. Dit wordt enigszins gecompenseerd door de oververtegenwoordiging van Limburgse scholen (7 van de 30) in de steekproef van het basisonderwijs.

Tabel 3.11
Spreading provincies in de definitieve steekproef secundair onderwijs

Provincie	N
Antwerpen	10
Oost-Vlaanderen	7
West-Vlaanderen	8
Limburg	0
Vlaams-Brabant	5

Onderstaande tabel biedt een samenvattend overzicht van de beoogde en definitieve steekproef. In totaal zijn er 57 scholen die wensten deel te nemen. Uiteindelijk bleven er van de 30 scholen voor basisonderwijs 27 scholen over. Voor het secundair onderwijs bestaat de steekproef uiteindelijk uit 30 scholen.

Tabel 3.12
Overzicht beoogde en definitieve steekproef voor basis en secundair onderwijs

	Beoogde steekproef	Definitieve steekproef
Basisonderwijs	30	27
Secundair onderwijs	51	30
Totaal	81	57

3. Instrumenten

Voor de dataverzameling werden twee vragenlijsten ontwikkeld: een uitgebreide versie voor de leerkrachten en een kortere versie voor de directie waarbij voornamelijk gepeild werd naar schoolkenmerken.

3.1. Vragenlijst leerkrachten

De vragenlijst voor leerkrachten (bijlage 1) start met het bevragen van **personalia**-gegevens zoals geslacht, leeftijd, SES, afkomst, vooropleiding, ervaring, huidige functie (voornaamste vak) en lesopdracht (in welk jaar/graad en welk(e) onderwijsvorm(en) en onderwijsniveaus(s)).

De tweede rubriek peilt naar de **opvattingen** van leerkrachten. Meer specifiek gaat het om de opvattingen omtrent onderwijs in het algemeen, hun opvattingen met betrekking tot diversiteit in onderwijs en hun opvattingen rond evaluatie.

In een derde onderdeel werden vragen gesteld met betrekking tot hun **ervaring van efficiëntie**. Zo werd gepeild naar hun doelmatigheidsbeleving en de mate waarin ze erin slagen om oog te hebben voor diversiteit in hun klas/onder leerlingen.

De vierde en tevens ook meest omvangrijke rubriek peilt naar de **klaspraktijk** van de leerkracht. Hierin wordt bevraagd in welke mate ze aan differentiatie doen in het algemeen. Meer specifiek wordt ook gepeild naar hoe hun evaluatiepraktijk eruit ziet inzake eerlijkheid, transparantie, verantwoording, betrokkenheid en impact. *Eerlijkheid* verwijst naar de mate waarin men ervoor zorgt dat de evaluatie fair is voor alle leerlingen zonder dat niet-relevante kenmerken van de leerlingen een effect hebben op het eindresultaat. *Transparantie* verwijst naar de mate waarin het voor de leerlingen duidelijk is hoe er geëvalueerd wordt en welke criteria beoordeeld zullen worden. Via de vragen die peilen naar *verantwoording* wordt nagegaan in welke mate de beoordeling gerechtvaardigd kan worden ten aanzien van de betrokkenen. De vragen die verwijzen naar *betrokkenheid* gaan na in welke mate de leerlingen deelnemen in het evaluatieproces. Het onderdeel rond *impact* gaat na in welke mate de evaluatiepraktijk invloed heeft op het leren van de leerling. Verder peilt deze rubriek ook naar hoe de evaluatiepraktijk van leerkrachten zich verhoudt tot de onderwijspraktijk. Of nog anders gezegd: de mate waarin evaluatie geïntegreerd is in de ruimere onderwijspraktijk. Ten slotte peilt deze rubriek naar de mate waarin leerkrachten zichzelf professionaliseren inzake evaluatie.

In een laatste rubriek wordt gepeild naar de school en meer specifiek naar haar **beleidsvoerend vermogen** omtrent evaluatie. Hierbij worden volgende indicatoren bevraagd: de gezamenlijke doelgerichtheid, de doeltreffende communicatie en het reflecterend vermogen.

De vragenlijst wordt afgerond met een open vraag waarbij respondenten eventuele bedenkingen of opmerkingen kunnen meedelen.

3.2. Vragenlijst directies

De vragenlijst voor directies (bijlage 2) start eveneens met een rubriek '**personalia**' waarin volgende zaken bevraagd worden: geslacht, leeftijd, vooropleiding, ervaring, specificaties met betrekking tot huidige functie (in welk onderwijsniveau en/of graad en welk(e) onderwijsvorm(en)).

Een tweede rubriek peilt naar de **schoolorganisatie** en specifiek naar hoe evaluatie georganiseerd wordt. Zo wordt gevraagd naar de wijze van groepering van leerlingen, of er al dan niet (her)examen- of proefwerkenperiodes zijn en voor welke groep(en) van leerlingen deze georganiseerd worden, welk leerlingvolgsysteem gebruikt worden. In deze rubriek wordt ook gevraagd naar de frequenties van klassenraden, oudercontacten en leerlingcontacten om de studievoortgang van leerlingen te bespreken. Een laatste set van vragen heeft betrekking op het rapport dat de school hanteert.

In een derde onderdeel worden enkele **schoolgegevens** opgevraagd, zoals het studieaanbod (graden en onderwijsvormen) van de school en het aantal leerkrachten dat er les geeft.

De laatste rubriek peilt, naar analogie met de vragenlijst voor leerkrachten, naar het **beleidsvoerend vermogen** van de school omtrent evaluatie. Hierbij worden dezelfde indicatoren bevraagd bij de directie: de gezamenlijke doelgerichtheid, de doeltreffende communicatie en het reflecterend vermogen.

Op het einde van de vragenlijst is er ook voor de directies een open vraag voorzien waarbij men eventuele bedenkingen of opmerkingen kan meedelen.

3.3. Ontwikkeling instrumenten

Voor de vragenlijsten werden enerzijds items zelf ontworpen en anderzijds werd er beroep gedaan op bepaalde schalen en items (die, indien nodig, vertaald werden naar de Vlaamse context). In bijlage 3 worden de details met betrekking tot deze bestaande schalen weergegeven. Eveneens wordt hierin de koppeling gemaakt met de opgenomen vragen en de variabelen zoals ze weergegeven worden in het gebruikte CIPO-model.

Een eerste versie van de vragenlijsten werd bezorgd aan de leden van de opvolgingsgroep op 21 maart 2019. Via een schriftelijke procedure bezorgden de leden van de opvolgingsgroep hun feedback op 4 april 2019. Aan de hand van deze feedback werden de vragenlijsten bijgewerkt.

Voor de afname van de survey werden de vragenlijsten ingevoerd in *LimeSurvey* (versie 2.73.1). Dit is een vrije, open source webapplicatie voor het uitwerken van online enquêtes. Vooraleer de link naar de online survey uitgestuurd werd naar de respondenten werd een pilot uitgevoerd om na te gaan of de survey technisch gebruiksklaar was.

4. Verloop dataverzameling

4.1. Opvolging dataverzameling

De link naar de vragenlijsten voor directies en leerkrachten werd na de paasvakantie (eind april) via mail verstuurd naar de contactpersoon van elke deelnemende school met de vraag om de survey in te vullen tegen eind mei. De contactpersonen bezorgden op hun beurt de link naar de vragenlijsten aan de directie en leerkrachten van het team met de bedoeling om zoveel mogelijk leerkrachten van het team te bereiken. De drie daaropvolgende weken ontvingen de scholen wekelijks een herinnering aan de deadline. In het herinneringsmiltje aan de contactpersoon werd voor elke school ook meegedeeld hoe hoog de respons inmiddels was.

Half mei bleek dat de respons nog te laag was wat leidde tot het langer open stellen van de survey. Enkele scholen hadden intussen ook laten weten dat juni beter zou uitkomen om de respons te verhogen, waardoor de deadline uitgesteld werd tot ongeveer het einde van het schooljaar. Naar analogie met de contactname tijdens het recruiteringsproces werd vanaf half mei ook ingezet op het telefonisch contacteren van de scholen om de respons te stimuleren. Deze telefonische contactname ging verder gepaard met het versturen van updates die de individuele scholen informeerden over de voorlopige respons op hun school.

Op 24 juni 2019 werd de online survey afgesloten.

4.2. Respons

We ontvingen respons van bijna alle 57 scholen die eerder bevestigd hadden om deel te nemen, met uitzondering van één school uit het secundair onderwijs waarvan er geen respons was voor de vragenlijst voor leerkrachten, noch voor de vragenlijst voor directie. Dit betekent dat de verdere analyses uitgevoerd werden op data van 56 scholen.

In totaal namen er 1120 leerkrachten deel aan de online bevraging, waarvan 419 leerkrachten in het basisonderwijs en 701 leerkrachten in het secundair onderwijs. De vragenlijst voor directies werd in totaal door 47 directies ingevuld waarbij 25 in het basisonderwijs en 22 in het secundair. Dit betekent dat er van de deelnemende scholen van het basisonderwijs twee scholen zijn waarvan er geen respons is van de directeur. In het secundair onderwijs zijn er acht scholen waarvan de vragenlijst voor directie niet beantwoord werd.

Tabel 3.13

Overzicht steekproeven en respons voor basis en secundair onderwijs

	Beogde steekproef	Steekproef na contactname	Gerealiseerde steekproef	Effectieve respons van individuen
Basisonderwijs	30	27	27	417 leerkrachten 25 directies
Secundair onderwijs	51	30	29	686 leerkrachten 22 directies
Totaal	81	57	56	1120 leerkrachten 47 directies

Hoofdstuk IV Univariate analyses

In dit hoofdstuk lichten we de univariate ofwel de beschrijvende analyses toe. Dit laat ons toe om de verzamelde data in een eerste fase te verkennen. De correlatiematrices die hierbij opgesteld werden voor de variabelen op leerkracht- en schoolniveau zijn raadpleegbaar in bijlage 4. Met het oog op verdere verklarende analyses waarbij we rekening willen houden met het feit dat de data multilevel van aard is, gingen we alvast voor enkele leerkrachtvariabelen na of het geoorloofd is om dit te aggregeren tot een schoolkenmerk.

Achtereenvolgens worden de resultaten van de beschrijvende analyses toegelicht met betrekking tot (1) de respondenten (leerkrachten en directies), ofwel de inputvariabelen volgens het CIPO-model, (2) de procesvariabelen en (3) de outputvariabelen. Aangezien deze survey zowel in basis- als in het secundair onderwijs afgenomen werd, worden de resultaten telkens afzonderlijk weergegeven per onderwijsniveau.

1. Beschrijving respondenten

1.1. Leerkrachten

In totaal hebben 1120 leerkrachten de vragenlijst ingevuld waarbij 417 van hen les geeft in het basisonderwijs en 686 in het secundair onderwijs. In het basisonderwijs is de meerderheid van de respondenten ($n=313$) werkzaam in het lager onderwijs, 83 werken in het kleuteronderwijs en 21 in de peuterklas.

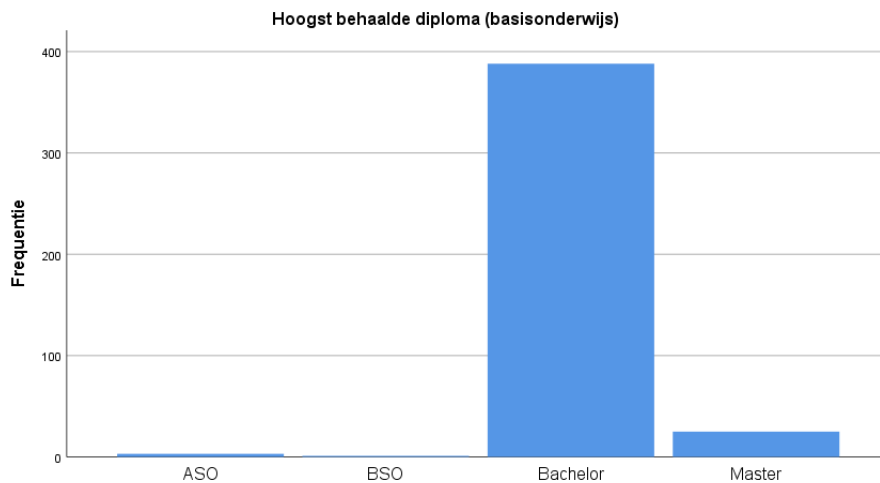
Met betrekking tot de variabele **geslacht** zien we dat over de onderwijsniveaus heen de meerderheid van de respondenten vrouw (77%) is. In het basisonderwijs zijn ongeveer negen op tien respondenten vrouw. De mannelijke leerkrachten zijn meer vertegenwoordigd in het secundair onderwijs: drie op de tien.

De gemiddelde **leeftijd** van alle bevroegde leerkrachten is 41 jaar. De verdeling voor leeftijd is in het basisonderwijs gelijkaardig aan het secundair onderwijs: in het basisonderwijs is de gemiddelde leeftijd ongeveer 40 jaar, in het secundair onderwijs 42 jaar. De minimum- en maximumleeftijd zijn eveneens ongeveer vergelijkbaar: de jongste leerkracht in het basisonderwijs is 22 jaar in het secundair 20 jaar. De oudste leerkracht in het lager is 63 jaar en in het secundair 67 jaar.

De bevroegde leerkrachten hebben gemiddeld genomen 16 jaar **ervaring** in het onderwijs. Naar analogie met de variabele 'leeftijd' is de verdeling met betrekking tot 'ervaring' in het basis- en secundair onderwijs gelijklopend. De minst ervaren leerkrachten in het basis- en secundair onderwijs hebben minder dan één jaar ervaring. De meest ervaren leerkracht heeft 46 jaar ervaring in het basisonderwijs en 41 jaar in het secundair onderwijs.

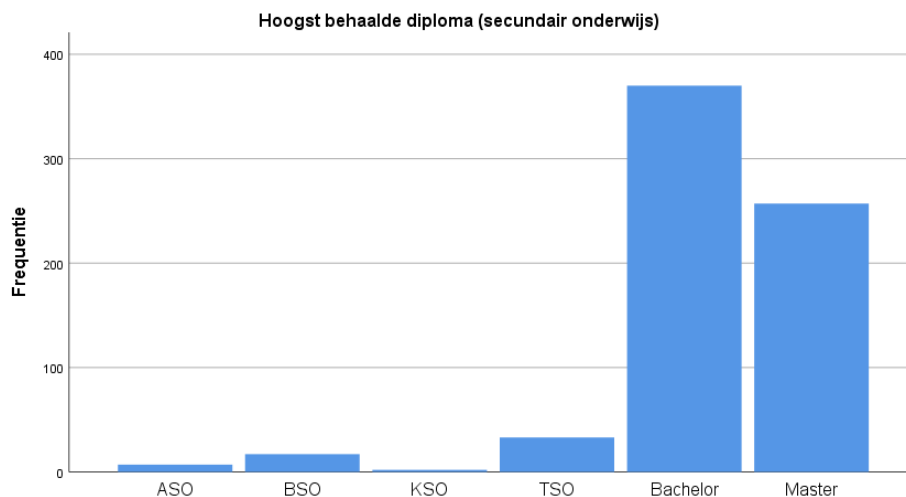
Met betrekking tot **vooropleiding** werd bij de respondenten enerzijds gepeild naar het hoogst behaalde diploma en anderzijds naar de wijze waarop men het bewijs van pedagogische bekwaamheid

verworven heeft. Het hoogst behaalde diploma bij de leerkrachten in het basisonderwijs is doorgaans een bachelordiploma (n=388). Vijfentwintig leerkrachten geven aan dat ze over een masterdiploma beschikken en bij vier respondenten verwijst het hoogst behaalde diploma naar het secundair onderwijs (waarbij drie van hen afstudeerden in het aso en één van hen in het bso).



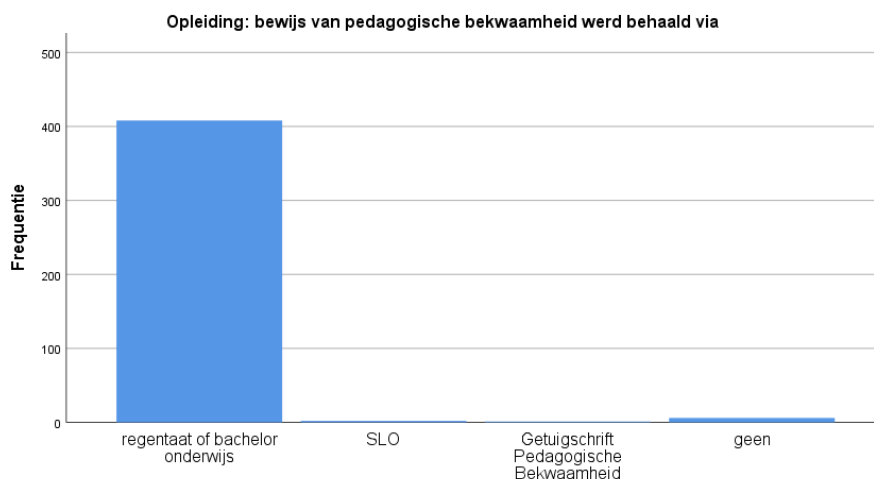
Grafiek 4.1: Hoogst behaalde diploma leerkrachten basisonderwijs

Bij de leerkrachten in het secundair onderwijs is het hoogst behaalde diploma meestal een bachelordiploma (n=370) of een masterdiploma (n=257). Negenenvijftig leerkrachten geven aan dat het hoogst behaalde diploma verwijst naar secundair onderwijs, meer specifiek gaat het om onderwijsvormen aso (n=7), bso (n=17), kso (n=2) en tso (n=33).



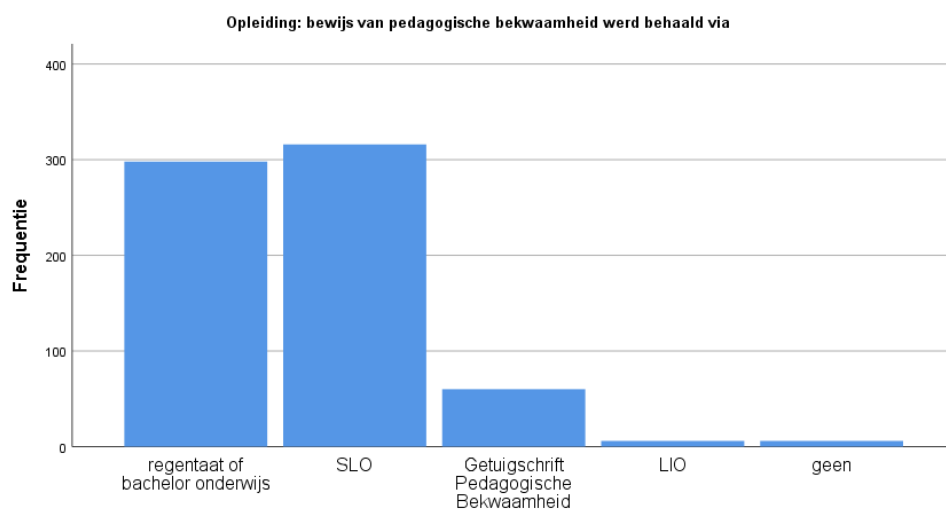
Grafiek 4.2: Hoogst behaalde diploma leerkrachten secundair onderwijs

In het basisonderwijs hebben bijna alle respondenten (n=408) een **bewijs van pedagogische bekwaamheid** behaald via een regentaats- of bacheloropleiding. Twee respondenten behaalden dit via een specifieke lerarenopleiding en één behaalde een Getuigschrift voor Pedagogische Bekwaamheid. Zes respondenten in het basisonderwijs beschikken niet over een bewijs van pedagogische bekwaamheid.



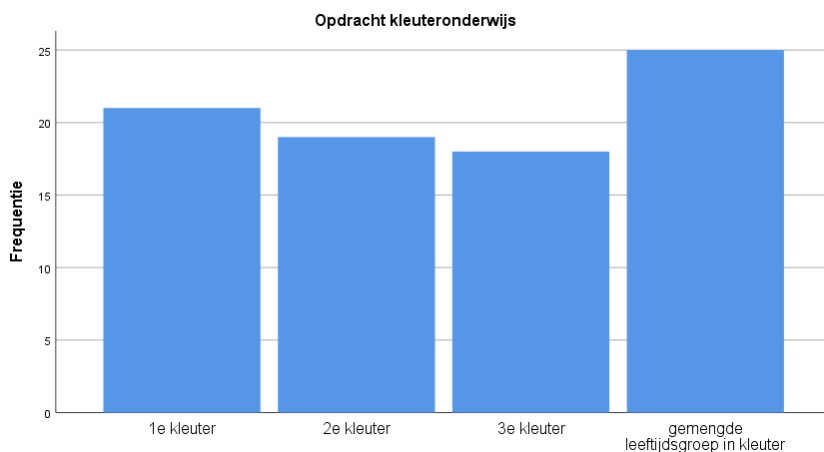
Grafiek 4.3: Vooropleiding voor bewijs van pedagogische bekwaamheid leerkrachten basisonderwijs

In het secundair onderwijs heeft de meerderheid van de leerkrachten zijn/haar bewijs van pedagogische bekwaamheid behaald via een specifieke lerarenopleiding (n=316) of via een regentaats- of bacheloropleiding (n=298). Zestig leerkrachten beschikken over een Getuigschrift Pedagogische Bekwaamheid. Zes leerkrachten zijn op het moment van de dataverzameling leraar-in-opleiding. Ten slotte zijn er zes leerkrachten in het secundair die niet over een bewijs van pedagogische bekwaamheid beschikken.



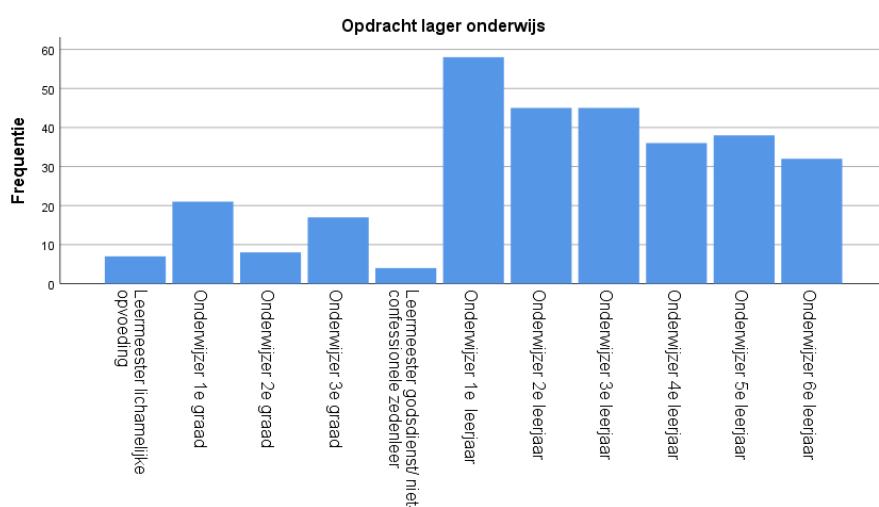
Grafiek 4.4: Vooropleiding voor bewijs van pedagogische bekwaamheid leerkrachten secundair onderwijs

In het basisonderwijs zien we dat de verschillende **leerjaren en/of leeftijdsgroepen** allemaal vertegenwoordigd worden door de respondenten. Onder de respondenten zijn er 21 die in het eerste kleuter werken, 19 in het tweede kleuter en 18 in de derde kleuterklas. Er zijn 25 leerkrachten die een kleuterklas hebben waar de leeftijden van de leerlingen gemengd zijn.



Grafiek 4.5: Opdracht leerkrachten in het kleuteronderwijs

In het lager onderwijs zijn er 58 leerkrachten van het eerste leerjaar, 45 in het tweede en het derde, 36 in het vierde, 38 in het vijfde, en 32 in het zesde leerjaar. Voor de scholen die niet georganiseerd zijn volgens het jaarklassensysteem konden leerkrachten aangeven in welke graad ze les geven: 21 leerkrachten in de eerste graad, acht in de tweede graad en 17 in de derde graad. Ten slotte waren er onder de respondenten ook leerkrachten die les geven aan meerdere leeftijden: leermeester lichamelijke opvoeding (n=7) en leermeester godsdienst/niet-confessionele zedenleer (n=4).



Grafiek 4.6: Opdracht leerkrachten in het lager onderwijs

De verdeling over de leerjaren/graden heen varieert sterker in het secundair onderwijs. Zo geven 265 respondenten aan dat hun huidige lesopdracht in de eerste graad is, 449 geven les in de tweede graad, 380 in de derde graad en 118 respondenten in de vierde graad of het 7^e jaar bso. Let wel dat de lesopdrachten van leerkrachten vaak betrekking hebben op meerdere graden waardoor één respondent bij deze vraag meerdere keren meegeteld kon worden.

Tabel 4.1

Overzicht aantal leerkrachten dat les geeft in de graden van het secundair

Graad	Aantal
1e	265
2e	449
3e	380
4e	118

Wat betreft de **onderwijsvormen** waarin de respondenten actief zijn op het moment van de dataverzameling, stellen we vast dat in de eerste graad 181 leerkrachten les geven in de a-stroom en 138 leerkrachten in de b-stroom. Met betrekking tot de overige onderwijsvormen zijn er 206 leerkrachten actief in het aso, 293 in het bso, 22 in het kso en 336 in het tso. Ook hier is het zo dat leerkrachten in meerdere onderwijsvormen kunnen les geven.

Tabel 4.2

Overzicht aantal leerkrachten dat les geeft in de onderwijsvormen

Onderwijsvorm	aantal
1a	181
1b	138
aso	206
bso	293
kso	22
tso	336

Met betrekking tot de **vakken** die leerkrachten in het secundair onderwijs (zie tabel 4.3), geldt eveneens dat leerkrachten met een lesopdracht in twee of meer vakken, meerdere keren meegeteld werden.

Tabel 4.3

Overzicht aantal leerkrachten dat les geeft in verschillende vakken van het secundair

Vak	Aantal
Nederlands	87
Klassieke talen	10
Moderne talen	102
Wiskunde en wetenschappen	145
Economie	30
Humane vakken (bijvoorbeeld geschiedenis, maatschappelijke vorming, cultuur- en gedragswetenschappen, ...)	53
PAV	56
Levensbeschouwelijke vakken	44
Plastische/muzikale opvoeding	39
Technische vakken (bijvoorbeeld elektromechanica, bouw, hout, esthetiek, zorg, kunst, handel)	125
Praktijkvakken (bijvoorbeeld elektromechanica, bouw, hout, esthetiek, zorg, kunst, handel)	136

Naast de lesopdracht gaven de leerkrachten aan in welke andere **functie(s)** ze ook ervaring hebben. Ongeveer drie vierde van de respondenten beschikt over bijkomende ervaring naast de lesopdracht. De meest voorkomende functies hierbij zijn klastitularis (in het secundair), zorgleerkracht, GOK/SES-leerkracht en leerkracht anderstalige nieuwkomers.

Tabel 4.4

Overzicht aantal leerkrachten dat ervaring heeft in andere functies in basis & secundair onderwijs

Functie	n (basisonderwijs)	n (secundair onderwijs)
Directie	19	14
Zorgcoördinator	54	11
Graadcoördinator	3	35
Leerkracht anderstalige nieuwkomer	69	87
GOK/SES-leerkracht	85	54
GOK/SES-coördinator	6	16
Zorgleerkracht	153	64
Klastitularis	N.v.t.	465
Andere ondersteunende functie	16	87
Nooit andere functie opgenomen	160	127

1.2. Directies

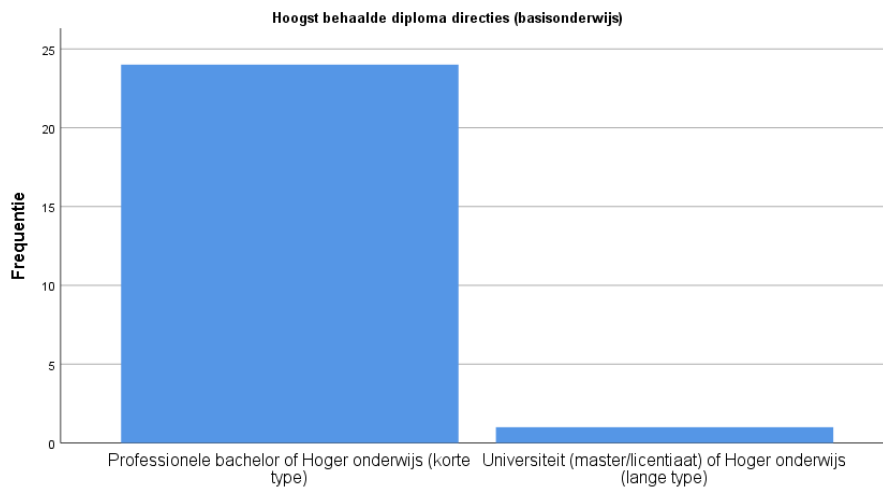
In totaal hebben 47 directies de vragenlijst ingevuld waarvan 25 van hen directeur zijn in een basisschool en 22 in een school voor secundair onderwijs. Bij de directies in het basisonderwijs zien we dat 18 van hen directeur zijn van een basisschool die zowel een aanbod heeft voor kleuter als lager onderwijs en zeven van hen zijn directeur in een afdeling voor lager onderwijs.

Met betrekking tot de variabele **geslacht** zien we dat over het algemeen de meerderheid van de scholen een vrouwelijke directie heeft. In het basisonderwijs is ongeveer drie op vier directies een vrouw. De mannen zijn, net zoals onder de leerkrachten, meer vertegenwoordigd in het secundair onderwijs, waar bijna de helft van de directies man is.

De gemiddelde **leeftijd** van alle bevroegde directies is ongeveer 50 jaar. De jongste directeur is 32 jaar, de oudste is 67 jaar. De verdeling van de leeftijd van directies is gelijkaardig voor beide onderwijsniveaus.

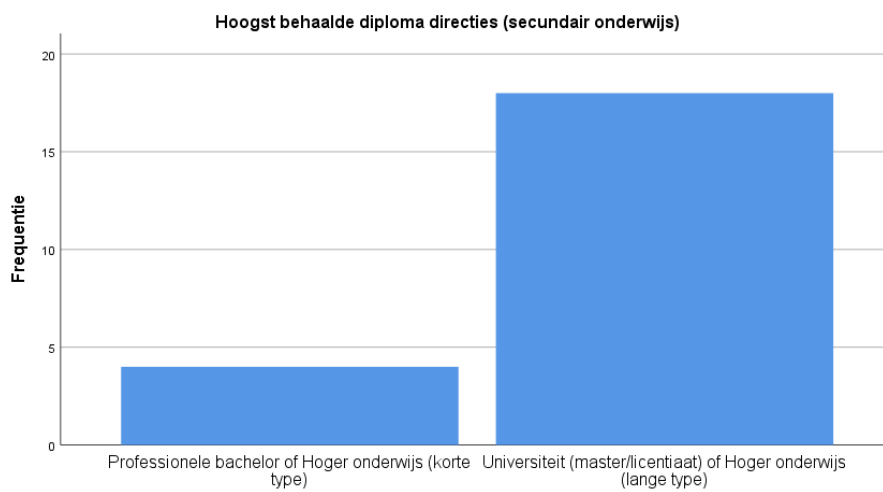
De gemiddelde **ervaring** als directie is ongeveer 11 jaar in het basisonderwijs waarbij dit varieert van minimum één jaar tot maximaal 24 jaar. In het secundair is de gemiddelde ervaring ongeveer zeven jaar waarbij de minst ervaren directie één jaar ervaring heeft en de meest ervaren 16 jaar directeur is. Naast de ervaring in de functie van directie gaven de respondenten ook aan hoeveel jaar ervaring ze hebben als leerkracht. De directies in het basisonderwijs hebben gemiddeld bijna 24 jaar ervaring (minstens vijf en hoogstens 42 jaar). In het secundair onderwijs hebben de bevroegde directies gemiddeld bijna 20 jaar ervaring, waarbij dit varieert van geen leservaring tot 34 jaar ervaring als leerkracht.

Met betrekking tot **vooropleiding** zien we dat acht op de tien directies in het secundair onderwijs over een master of licentiaatsdiploma beschikken als hoogste diploma.



Grafiek 4.7: Hoogst behaalde diploma directies basisonderwijs

De directies in het secundair onderwijs zijn hoger opgeleid dan de bevroagde directies in het basisonderwijs waar ze allen, op één na, een bacheloropleiding of hoger onderwijs korte type als hoogste opleidingsniveau hebben.



Grafiek 4.8: Hoogst behaalde diploma directies secundair onderwijs

2. Beschrijving procesvariabelen

2.1. Differentiatiepraktijk leerkrachten

De differentiatiepraktijk van leerkrachten geeft aan in welke mate leerkrachten in hun klaspraktijk rekening houden met verschillen tussen leerlingen. Hoewel één van de opgenomen items betrekking heeft op de evaluatiepraktijk (*ik pas mijn evaluatievorm aan in functie van bepaalde (groepen van) leerlingen*), wordt hier voornamelijk de differentiatiepraktijk in het algemeen bedoeld. Zo zijn ook items als *'ik bied leerinhouden aan afgestemd op het leerniveau van elke leerling in mijn klas'* opgenomen. De items werden gescoord op een Likertschaal met zeven punten waarbij nul aangeeft dat ze dit nooit doen en zes dat ze dit altijd doen.

In het basisonderwijs zien we dat de leerkrachten gemiddeld 3,41 scoren, wat erop wijst dat ze regelmatig tot vaak differentiëren in hun klaspraktijk. In het secundair onderwijs wordt er iets minder gedifferentieerd door leerkrachten: de gemiddelde score is er 2,74.

2.2. Aandacht voor de kwaliteitscriteria impact, betrokkenheid, eerlijkheid, transparantie en verantwoording

Wat de evaluatiepraktijk van leerkrachten betreft werd er ook gepeild naar de mate waarin men aandacht heeft voor de verschillende kwaliteitscriteria: impact, betrokkenheid van de leerling, eerlijkheid, transparantie en verantwoording. De respondenten konden dit aangeven op een Likertschaal waarbij nul staat voor 'helemaal oneens' en vier staat voor 'helemaal eens'.

Het criterium '**impact**' verwijst naar het effect dat evaluatie heeft op het uiteindelijke leergedrag van leerlingen. Hiertoe werden onder andere volgende items gebruikt: *'mijn evaluatietaken dagen de leerlingen uit om datgene wat ze geleerd hebben in alle breedte te tonen, mijn evaluatietaken zorgen ervoor dat leerlingen zich voorbereiden door grondig en diepgaand te studeren'*. De bevraagde leerkrachten in het secundair onderwijs rapporteren een iets sterkere impact (gemiddeld 2,95) van hun evaluatiepraktijk dan in het basisonderwijs (gemiddeld 2,72).

De schaal die de '**betrokkenheid**' bij evaluatie meet, peilt naar de mate waarin de leerling actief betrokken wordt bij de evaluatie. De opgenomen items verwijzen naar het gebruik van peer- en zelfevaluatie, maar ook naar de bijdrage van leerlingen in de voorbereidende fasen van evaluatie, zoals bijvoorbeeld: *'de leerlingen hebben inspraak over de manier waarop er geëvalueerd zal worden'* of *'ik laat de leerlingen mee nadenken over de criteria waaraan hun prestaties moeten voldoen.'* De bevraagde leerkrachten nemen een eerder neutraal standpunt in, waarbij de betrokkenheid van de leerlingen in het basisonderwijs (gemiddeld 2,30) iets meer aandacht krijgt dan in het secundair onderwijs (gemiddeld 2,21).

De items die samen het criterium '**eerlijkheid**' vormen, peilen naar de mate waarin de leerkracht aandacht heeft voor het feit dat evaluatie *billijk* ofwel *fair* moet zijn voor alle leerlingen. Dit houdt in dat niet-relevante kenmerken van een leerling (zoals bijvoorbeeld leerstoornissen of een thuistaal anders dan het Nederlands) geen invloed hebben op het resultaat. Voorbeelden van de opgenomen items hierbij zijn: *'ik zorg ervoor dat de evaluatietaken ook duidelijk zijn voor taalzwakke leerlingen'* of *'ik ben even streng voor leerlingen die makkelijk leren en meestal erg goede resultaten behalen als voor leerlingen die hun best doen maar meer moeite hebben met studeren.'* De gemiddelde scores over de

onderwijsniveaus heen laten zien dat leerkrachten in het secundair onderwijs (gemiddeld 3,09) hier meer aandacht aan besteden dan de leerkrachten in het basisonderwijs (gemiddeld 2,70).

De **'transparantie'** verwijst naar de mate waarin de leerlingen geïnformeerd worden over de evaluatie, zoals waarover (*ik licht mijn leerlingen op voorhand in over de leerdoelen die ik beoordeel in de evaluatietaken*), wanneer (*ik plan nooit onverwachte evaluaties waarop de leerlingen zich niet kunnen voorbereiden*) en op basis van welke criteria (*ik bezorg de leerlingen op voorhand alle informatie over de evaluatietaken zodat ze zich optimaal kunnen voorbereiden*) men geëvalueerd wordt. De bevroegde leerkrachten geven aan dat ze in hun evaluatiepraktijk hierover communiceren naar hun leerlingen toe. De leerkrachten in het secundair onderwijs (gemiddeld 3,42) besteden hier iets meer aandacht aan dan die in het basisonderwijs (gemiddeld 3,09).

Het **'verantwoordingsaspect'** binnen de evaluatiepraktijk verwijst naar de mate waarin de evaluatie gerechtvaardigd kan worden, of anders gezegd: of het voor de betrokkenen helder is hoe een bepaald resultaat tot stand kwam. Onder andere volgende items werden hiervoor beoordeeld door de respondenten: *'ik maak gebruik van een scoringsformulier voor het beoordelen van evaluatietaken van leerlingen'* en *'voor de afname is het voor mijzelf reeds duidelijk wat leerlingen moeten kunnen om het maximum van de punten te behalen'*. Volgens de respondenten is dit het kwaliteitscriterium dat het meeste aandacht krijgt in hun evaluatiepraktijk, waarbij dit bij de leerkrachten in het secundair sterker aanwezig is (gemiddeld 3,51) dan in het basisonderwijs (gemiddeld 3,32).

De resultaten laten zien dat alle kwaliteitsaspecten aandacht krijgen in de evaluatiepraktijk van leerkrachten. De gemiddeldes voor de verschillende schalen situeren zich namelijk telkens aan het positieve spectrum. Dit zou mogelijk te maken kunnen hebben met effect van sociale wenselijkheid bij het invullen van de bevraging. Bij het onderling vergelijken van de verschillende criteria zien we dat de kwaliteitsaspecten 'verantwoording' en 'transparantie' duidelijk meer aandacht krijgen in de evaluatiepraktijk van leerkrachten dan 'eerlijkheid' en 'impact'. De betrokkenheid van leerlingen bij het evaluatieproces krijgt het minste aandacht. Over het algemeen kunnen we stellen dat de leerkrachten in het secundair onderwijs iets meer begaan zijn met de verschillende kwaliteitscriteria dan de leerkrachten in het basisonderwijs, met uitzondering van 'betrokkenheid' dat in het basisonderwijs een iets hogere gemiddelde score behaalt dan in het secundair onderwijs.

Op schoolniveau zien we dat er enkel sprake is van een gedeelde praktijk onder leerkrachten met betrekking tot de aspecten 'eerlijkheid' (ICC = 0,73) en 'transparantie' (ICC = 0,66). De intraclasscorrelaties voor de overige kwaliteitsaspecten zijn te laag waardoor we niet kunnen zeggen dat er sprake is van gedeelde praktijken onder leerkrachten. Bijgevolg is het niet aangewezen om de aspecten 'verantwoording', 'impact' en 'betrokkenheid' als een kenmerk van de school te beschouwen.

2.3. Professionalisering van leerkrachten

De schaal 'professionalisering' geeft weer in welke mate leerkrachten hun kennis en vaardigheden omtrent het evalueren van leerlingen bijsturen. Dit wordt nagegaan aan de hand van items zoals 'ik bespreek problemen met mijn evaluatiepraktijk met anderen om zo bij te leren' en 'ik lees vakliteratuur over het evalueren van leerlingen.' De items werden gescoord op een Likertschaal van nul (bijna nooit) tot drie (bijna altijd). De leerkrachten in het basisonderwijs geven over het algemeen aan dat ze soms inzetten op professionalisering inzake 'evaluatie van leerkrachten' (gemiddeld 1,30). De leerkrachten in het secundair onderwijs zetten iets meer in op professionalisering voor hun evaluatiepraktijk (gemiddeld 1,48).

De intraclasscorrelatie (ICC = 0,61) overschrijdt de drempelwaarde van 0,60 wat ons toelaat om de mate waarin leerkrachten inzetten op professionalisering te aggregeren op schoolniveau door het gemiddelde te nemen per school (Glick, 1985). Dit betekent dat het professionaliseringsbeleid van een school inzake evaluatie als een kenmerk van de school beschouwd kan worden. Bijgevolg zullen we in de verdere analyses deze geaggregeerde maat opnemen als een schoolvariabele.

2.4. Opvattingen van leerkrachten met betrekking tot onderwijs

De onderwijsopvattingen van leerkrachten werden bevraagd aan de hand van stellingen die gescoord konden worden op een Likertschaal waarbij nul staat voor helemaal oneens en vier voor helemaal eens. De schaal voor onderwijsopvattingen omvat twee dimensies die verwijzen naar de mate van leerlinggerichtheid en de mate van leerstofgerichtheid. Deze laatste dimensie verwijst naar het belang dat men besteedt aan beroepsvoorbereiding, discipline en conformiteit van de leerlingen. Deze dimensie sluit aan bij een productgerichte onderwijsbenadering met nadruk op het leveren van prestaties en behalen van goede cijfers. De leerlinggerichte dimensie stelt de vorming van de lerende meer centraal, waarbij de lerende ook mee inspraak krijgt en waarbij de nadruk ligt op het proces van leren en het leren van elkaar.

Tabel 4.5

Overzicht variabelen voor opvattingen van leerkrachten met betrekking tot onderwijs

Variabele	Items	Cronbach's alpha
Opvatting onderwijs: leerlinggericht	13	0,843
Opvatting onderwijs: leerstofgericht	12	0,837

De gemiddelde scores van de leerkrachten laten zien dat de leerlinggerichte opvattingen sterker aanwezig zijn dan de leerstofgerichte opvattingen, dit voor zowel het basisonderwijs als het secundair onderwijs. De gemiddelde scores in het basisonderwijs zijn respectievelijk 3,20 en 2,51; in het secundair onderwijs 3,19 en 2,64.

Binnen dit onderzoek is het interessant om te weten of de opvattingen van leerkrachten gedeeld worden op niveau van de school. Indien dit het geval is, kunnen we deze maat op organisatieniveau aggregeren en meenemen als schoolvariabele in de verdere, verklarende analyses. Om te bepalen of overtuigingen gedeeld zijn binnen een school, wordt de Spearman-Brown-formule gehanteerd (Glick, 1985). De interpretatie van deze formule stelt dat de intraclasscorrelatie (ICC), gemeten op basis van

een ANOVA via de formule ($\text{BetweenMeanSquare} - \text{WithinMeanSquare} / \text{BetweenMeanSquare}$), groter moet zijn dan 0,60 om aggregatie door het nemen van het gemiddelde per school toe te laten. De intraclasscorrelaties voor de onderwijsopvattingen, respectievelijk 0,28 en 0,51 voor leerlinggericht en leerstofgericht, zijn te laag om te kunnen aggregeren op schoolniveau. Dit betekent dat er geen sprake is van gedeelde overtuigingen of een cultuur op schoolniveau met betrekking tot de onderwijsopvattingen.

2.5. Opvattingen van leerkrachten met betrekking tot diversiteit

Meer specifiek peilden de items inzake de diversiteitsopvatting naar de mate waarin leerkrachten het omgaan met diversiteit als een kerntaak van de leerkracht en het onderwijs beschouwen. Het gaat hierbij om een algemene visie op de maatschappelijke verantwoordelijkheid die het onderwijs heeft ten aanzien van het leren van kwetsbare leerlingen. Leerkrachten die ervan overtuigd zijn dat dit tot hun kerntaak behoort, zullen namelijk sneller geneigd zijn om hun leeromgeving zo in te richten dat ze maximale leer- en ontwikkelingskansen creëren voor alle leerlingen. De opvattingen van leerkrachten in het basisonderwijs zijn niet zo verschillend van die van leerkrachten in het secundair onderwijs. Beide groepen scoren gemiddeld respectievelijk 3,84 en 3,71 op een zevenpuntschaal, wat erop wijst dat leerkrachten een neutrale tot licht positieve houding hebben ten aanzien van de opdracht die onderwijs heeft inzake diversiteit.

De intraclasscorrelatie ($ICC = 0,69$) overschrijdt de drempelwaarde van 0,60 wat ons toelaat om de opvattingen van leerkrachten met betrekking tot diversiteit te aggregeren op schoolniveau door het gemiddelde te nemen per school (Glick, 1985). Dit betekent dat de overtuigingen van leerkrachten met betrekking tot diversiteit gedeeld worden binnen scholen. Bijgevolg zullen we in de verdere analyses deze geaggregeerde maat opnemen als een schoolvariabele.

2.6. Bekwaamheidsgevoel van leerkrachten met betrekking tot oog hebben voor diversiteit

In tegenstelling tot de voorgaande schaal, heeft deze schaal niet zozeer betrekking op de opvattingen, maar eerder op het bekwaamheidsgevoel van de leerkracht. De items in deze schaal gaan na in welke mate men zichzelf bekwaam voelt om oog te hebben voor de verschillende facetten van diversiteit bij leerlingen. Meer specifiek zijn hier items opgenomen zoals *'tijd maken voor diepgaande individuele gesprekken met al uw leerlingen om hen beter te ondersteunen'* of *'een zicht te krijgen op de leerstijl van de leerling, door bewust te kijken hoe hij/zij reageert op werkvormen en groepeeringsmethoden'*. Op een negenpuntschaal geven de leerkrachten van het basis- en secundair onderwijs ongeveer een gelijke score die erop wijst dat ze zich hier een beetje tot redelijk bekwaam in voelen (respectievelijk gemiddeld 5,53 en 5,08).

2.7. Opvattingen van leerkrachten met betrekking tot evaluatie

Om een zicht te krijgen op de opvattingen die leerkrachten erop nahouden als het gaat over evaluatie, werden een 27-tal items opgenomen die verwijzen naar mogelijke opvattingen. Om de opvattingen in kaart te brengen werd een factoranalyse uitgevoerd op deze 27 items. Dit resulteert in 6 factoren die samen 57,7% van de totale variantie verklaren. Drie factoren omvatten echter slechts één of twee items. Vervolgens werd een principale componentenanalyse uitgevoerd waarbij het aantal

componenten gereduceerd werd tot drie. Dit leverde de beste *fit* op voor de data en verklaarde 44,23% van de totale variantie. Binnen de drie componenten werden de items met een lading groter dan 0,420 behouden voor het samenstellen van de variabele.

Op basis van de principale componentenanalyse onderscheiden we drie componenten voor de opvattingen van leerkrachten met betrekking tot evaluatie. Een eerste variabele verwijst naar de opvatting dat evaluatie voornamelijk dient om te **informer**en. Wanneer we de 17 items onder de loep nemen, dan kan dit informerende karakter verwijzen naar de leerkracht (*'informatie verkregen door evaluatie leidt tot aanpassingen in de manier van lesgeven'*), de school (*'evaluatie biedt informatie over hoe goed scholen het doen'*) of de leerling zelf (*'evaluatie biedt leerlingen feedback over hun prestaties'*). Bovendien wordt de opvatting dat evaluatie een informerend doel dient, positief gewaardeerd, zo blijkt uit de positieve lading van volgende items: *'toetsresultaten zijn consistent'* en *'op toetsresultaten kan vertrouwd worden'*. Een tweede component (met negen items) verwijst naar een kritische houding ten aanzien van evaluatie. De items hier verhullen enerzijds een negatieve houding ten aanzien van de waarde van toetsresultaten (onnauwkeurig, meetfouten, oneerlijk, ...). Anderzijds verwijzen de items naar een beperkt gebruik van toetsresultaten: *'leerkrachten doen aan evaluatie maar maken weinig gebruik van de resultaten ervan'* en *'toetsresultaten worden geklasseerd en genegeerd'*. Een laatste component (drie items) verwijst naar de opvatting dat evaluatie **geen impact op instructie** heeft: *'evaluatie heeft weinig effect op de manier van lesgeven'*. De overige twee items crossladen hier met het item dat de opvatting van 'informer

en' vat, maar dan negatief: *'informatie verkregen door evaluatie leidt tot aanpassingen in de manier van lesgeven'* en *'evaluatie zorgt ervoor dat leerlingen aangepaste instructie kunnen krijgen'*.

De Cronbach's alpha's van de subschalen laten zien dat enkel de variabele 'opvattingen evaluatie informeren' en 'opvattingen evaluatie kritisch' de drempelwaarde van 0,70 halen. De subschaal 'opvattingen evaluatie: geen impact op instructie' levert een lagere Cronbach's alpha op, maar aangezien dit slechts drie items omvat, achten we dit als voldoende hoog om dit als een valide en betrouwbare variabele op te nemen.

Tabel 4.6

Overzicht variabelen voor opvattingen van leerkrachten met betrekking tot evaluatie

Variabele	Items	Cronbach's alpha
Opvattingen evaluatie: informeren	17	0,906
Opvattingen evaluatie: kritisch	9	0,742
Opvattingen evaluatie: geen impact op instructie	3	0,569

De respondenten konden de stellingen scoren op een positief georiënteerde zespuntenschaal van nul tot vijf waarbij nul staat voor 'helemaal oneens', één voor 'grotendeels oneens', twee voor 'beetje eens', drie voor 'gemiddeld eens', vier voor 'grotendeels eens' en vijf voor 'helemaal eens'. De gemiddelde scores van de leerkrachten laten zien dat de leerkrachten in het basisonderwijs overtuigd zijn van de positieve waarde van evaluatie (gemiddeld 2,98), namelijk dat dit zowel voor de leerkracht, de school als de leerling zelf nuttige informatie oplevert. De kritische houding ten aanzien van evaluatie scoort gemiddeld één punt lager (gemiddeld 1,99), wat inhoudt dat de leerkrachten hier een beetje mee akkoord gaan, maar in mindere mate dan met de positieve houding ten aanzien van evaluatie. In het secundair onderwijs is er een gelijkaardige trend vast te stellen, al liggen de opvattingen hier iets verder uit elkaar. De bevraagde leerkrachten in het secundair onderwijs scoren gemiddeld 3,12 voor

de positieve opvatting die verwijst naar het informierend karakter van evaluatie en 1,89 op de eerder kritische houding ten aanzien van evaluatie.

De intraclasscorrelaties voor elk van de subschalen (ICC informeren = 0,37; ICC kritisch = 0,41 en ICC geen impact op instructie = 0,37) laten zien dat er voor de onderzochte scholen geen sprake is van een cultuur omtrent de opvattingen rond evaluatie. De overtuigingen van leerkrachten met betrekking tot evaluatie variëren binnen eenzelfde school wat ons ervan weerhoudt om dit als een kenmerk van een school te beschouwen.

2.8. Doelmatigheidsbeleving leerkrachten

De vragenlijst voor leerkrachten peilde naar hun doelmatigheidsbeleving (*self-efficacy*) ofwel de mate waarin de leerkrachten geloven in hun eigen competenties inzake instructiestrategieën, klasmanagement en leerlingenbetrokkenheid. Dit werd nagegaan aan de hand van items zoals *'in welke mate slaagt u erin om de leerlingen de regels te laten naleven in de klas'* en *'in welke mate slaagt u erin om leerlingen te motiveren die weinig interesse tonen in hun schoolwerk'*. Dit werd gemeten aan de hand van een Likertschaal waarbij nul staat voor 'niet' en acht voor 'veel'.

De gemiddelde scores per onderwijsniveau tonen aan dat de leerkrachten in het basisonderwijs (gemiddeld 6,18) zich iets competentier voelen dan de bevroegde groep in het secundair onderwijs (gemiddeld 5,94).

2.9. Organisatie van de school

In de directievragenlijst werden een aantal vragen opgenomen die de organisatie van de school in kaart brengen. Zo werd er gepeild of de scholen georganiseerd zijn volgens het jaarklassensysteem, of ze (her)examens organiseren en hoe frequent er klassenraden, oudercontacten en leerlingcontacten georganiseerd worden. Tot slot werd ook gepeild naar het gebruik van een leerlingenvolgsysteem en de manier waarop het rapport op school gebruikt wordt.

We dienen hier rekening te houden met het feit dat er voor enkele van de deelnemende scholen geen gegevens beschikbaar zijn over de organisatie wegens non-respons bij de directievragenlijst. De beschrijvende resultaten voor het basisonderwijs zijn gebaseerd op een aantal van 25 directies, deze van het secundair onderwijs op een aantal van 22 directies.

2.9.1. Jaarklassensysteem

In het basisonderwijs houdt de meerderheid (n=18) van de scholen vast aan het jaarklassensysteem om de leerlingen te groeperen. Zeven basisscholen geven aan dat ze hier soms van afwijken en voor bepaalde activiteiten jaarklasdoorbekend werken. In het secundair onderwijs is de verdeling omgekeerd: 13 van de 22 scholen geven aan dat ze soms jaarklasdoorbekend werken en negen scholen geven aan dat ze steeds het jaarklassensysteem hanteren.

2.9.2 (Her)examens

In het secundair onderwijs geven de directies die participeerden aan het onderzoek aan dat ze allen (n=22) gebruik maken van een systeem met examens en iets meer dan de helft (n=18) biedt ook

herexamens aan. In het basisonderwijs organiseert de meerderheid (19 van de 25) van de scholen examens, maar worden er geen herexamens ingericht.

2.9.3. Frequentie klassenraden, oudercontacten en leerlingcontacten

Volgens onderstaande tabel organiseren alle scholen van wie de directie de vragenlijst invulde, klassenraden en oudercontacten. In het basisonderwijs organiseert men **klassenraden** minstens twee keer per jaar tot hoogstens vijf of meerder keren per schooljaar. De meerderheid van de basisscholen organiseert dit drie keer per schooljaar. In het secundair onderwijs wordt dit frequenter ingepland: daar varieert de frequentie van minstens drie keer tot vijf of meer dan vijf keer, wat overigens het geval is voor de meerderheid van de scholen.

Met betrekking tot het **oudercontact** varieert de frequentie van minstens twee keer per schooljaar tot vijf of meerdere keren per schooljaar. De meest voorkomende frequentie verschilt naargelang het onderwijsniveau: in het basisonderwijs is dit drie keer per schooljaar en in het secundair vier keer.

In tegenstelling tot de klassenraden en oudercontacten zijn georganiseerde contacten met leerlingen via reflectiegesprekken of **leerlingcontacten** beperkter ingeburgerd. Over het basis- en secundair onderwijs zijn er acht scholen die dit niet doen en negen scholen die dit één keer per schooljaar inplannen. In meer dan de helft van de scholen (n=32) wordt dit twee of meerdere keren per schooljaar georganiseerd.

Tabel 4.7

Overzicht frequenties klassenraden, oudercontacten en leerlingcontacten in basis- en secundair onderwijs.

Frequentie per schooljaar	Klassenraad		Oudercontact		Leerlingcontact	
	Basis	Secundair	Basis	Secundair	Basis	Secundair
Niet	0	0	0	0	3	5
1	0	0	0	0	3	4
2	3	0	4	3	10	3
3	14	1	15	2	5	4
4	4	7	5	12	0	3
5 of meerdere keren	4	14	1	5	4	3

2.9.4. Leerlingenvolgsysteem

Alle scholen waarbij de directie de vragenlijst invulde (n=47), maken gebruik van een leerlingenvolgsysteem. Onderstaande tabel geeft weer op welke leerlingenvolgsystemen de scholen beroep doen. De secundaire scholen maken voornamelijk gebruik van 'Smartschool' en twee scholen verwezen ook naar 'Schoolware'. In het lager onderwijs is er slechts één school die 'Smartschool' hanteert voor het opvolgen van hun leerlingen. Over het algemeen zien we dat er in het lager onderwijs veel meer variatie is. 11 scholen maken gebruik van het leerlingenvolgsysteem van het Vrij CLB en onder de overige leerlingenvolgsystemen behoorden 'Schoolonline', 'Broeckx', 'loomniwize', 'Wisa', 'Questi' en 'Zorgkompas'. Eén school gaf ten slotte aan dat ze zelf een eigen systeem ontwikkeld hebben om de voortgang van hun leerlingen op te volgen.

Tabel 4.8

Overzicht van gerapporteerde leerlingvolgsystemen in het basis- en secundair onderwijs.

Onderwijsvorm	Leerlingvolgsysteem	Frequentie
Basisonderwijs	LVS van VCLB	11
	Schoolonline	4
	Broeckx	2
	Volg	2
	loomniwize	1
	Smartschool	1
	Wisa	1
	Questi	1
	Zorgkompas	1
	Volgsysteem ontwikkeld door de school	1
	Totaal	25
Secundair onderwijs	Schoolware	2
	Smartschool	20
	Totaal	22

2.9.5. Gebruik van het rapport

Onderstaande tabel geeft de gemiddelde scores van de items met betrekking tot het rapport weer voor zowel het basis- als het secundair onderwijs. Deze gemiddelden werden berekend op basis van de positionering op een vijfpuntenschaal waarbij nul staat voor 'helemaal oneens' en vijf voor 'helemaal eens'.

Wat de aard van het rapport betreft, zien we dat het cijfer- of letterrapport het meest gangbaar is in zowel het basis- als het secundair onderwijs. Het weergeven van doelen is over het algemeen in mindere mate van toepassing bij de rapportering, waarbij dit nog vaker voorkomt in het basisonderwijs. Een rapportering in de vorm van een groeiboek of portfolio komt slechts heel beperkt voor: in het secundair onderwijs in vier scholen en in het basisonderwijs in één school.

Het weergeven van een vergelijking met medeleerlingen aan de hand van een klasprofiel, gemiddelde of mediaan zien we in de helft van de scholen in het secundair onderwijs en in de andere helft helemaal niet. In het basisonderwijs is de verhouding anders: iets minder dan vier vijfde van de basisscholen neemt een dergelijke vergelijking op.

Wat de functie van het rapport betreft, zien we over het algemeen dat het rapport zich voornamelijk sterk richt op de communicatie naar ouders over de vooruitgang en prestaties van hun kind. In iets mindere mate wordt het rapport beschouwd als communicatiemiddel naar de leerling toe. Dit laatste is nog meer het geval in het secundair onderwijs. De mate waarin de bevroegde directies het rapport beschouwen als een manier om leerkrachten aan te zetten tot evalueren is sterk afhankelijk van school tot school. De frequenties geven een heel gevarieerd beeld voor beide onderwijsvormen waardoor de gemiddelden dicht bij het 'neutrale' standpunt aanleunen.

Tabel 4.9

Overzicht gemiddelde scores per item over het rapport in het basis- en secundair onderwijs.

Items	Basis	Secundair
Het rapport geeft cijfers/letters weer voor de meeste vakken.	3,24	3,91
Het rapport geeft geen klasprofiel, mediaan of klasgemiddelde weer.	2,96	2,05
Het rapport kan omschreven worden als een doelenrapport.	2,16	1,41
Het rapport besteedt geen aandacht aan het gedrag en de attitudes van leerlingen.	0,88	1,00
Op onze school gebruiken we geen klassiek 'rapport'. We gebruiken een groeiboek of portfolio waarin het leerproces van de leerling bijgehouden wordt.	1,04	0,95
Het rapport en wat erin opgenomen wordt, wordt op onze school frequent bijgestuurd door leerkrachten en/of directie.	2,40	2,64
Het rapport is een goede weergave van hoe wij naar onze leerlingen kijken.	2,68	2,55
Het rapport dient als instrument om leerkrachten aan te zetten tot het evalueren van de leerlingen.	2,24	2,18
Het rapport dient om de leerling te informeren over zijn/haar vooruitgang en prestaties.	2,60	3,27
Het rapport dient om de ouders te informeren over de vooruitgang en prestaties van hun kind.	3,16	3,36

2.9.6. Gebruik gevalideerde toetsen

De meerderheid van de bevraagde directies geeft aan dat ze op hun school gevalideerde of genormeerde toetsen gebruiken om na te gaan hoe ze het als school doen. Dit is het geval voor alle basisscholen waarvan de directie de vragenlijst invulde. In het secundair is dit het geval volgens 15 van de 22 bevraagde directies.

De bevraging (aan de hand van een gesloten 'ja-neenvraag') laat niet toe om zicht te krijgen op de aard van de toetsen. Bovendien liet het casestudieonderzoek (Ysenbaert et al., 2018) zien dat de wijze waarop gevalideerde toetsen aangewend worden sterk verschilt van school tot school. Dit varieert van het louter afnemen en registreren tot het aanwenden als instrument voor monitoring dat richtinggevend is voor beleidskeuzes van de school. Tussen beide uitersten waren er bij de onderzochte cases ook voorbeelden waarbij de resultaten van gevalideerde toetsen (gedeeltelijk) opgenomen werden als resultaat bij de rapporten van leerlingen.

2.10. Beleidsvoerend vermogen van de school inzake evaluatie

Zowel in de vragenlijst voor leerkrachten als voor directies werd gepeild naar drie indicatoren voor het beleidsvoerend vermogen van de school inzake het evaluatiebeleid: gezamenlijke doelgerichtheid, doeltreffende communicatie en reflecterend vermogen. Dit werd bevraagd aan de hand van een Likertschaal met vijf punten waarbij nul staat voor 'helemaal oneens' en vier voor 'helemaal eens'. Met betrekking tot de bevraging van de leerkrachten zien we dat men over het algemeen een beperkte mate van beleidsvoerend vermogen ervaart. In het basisonderwijs is dit nog iets sterker dan in het secundair onderwijs. De bevraging bij directies leverde een iets positievere inschatting op van het

beleidsvoerend vermogen: de gemiddelde scores zijn ongeveer 0,09 tot 0,47 punten hoger. Enkel voor ‘reflecterend vermogen’ komen de ervaringen van leerkrachten min of meer overeen met deze van de directies. In de verdere analyses nemen we voor ‘beleidsvoerend vermogen’ het perspectief van de leerkrachten in rekening.

Tabel 4.10.

Overzicht gemiddelde scores per onderwijsniveau met betrekking tot het beleidsvoerend vermogen

Onderwijsniveau	Basisonderwijs		Secundair onderwijs	
	Leerkrachten	Directies	Leerkrachten	Directies
Variabele				
Gezamenlijke doelgerichtheid	2,56	2,70	2,30	2,65
Doeltreffende communicatie	2,65	3,01	2,31	2,78
Reflecterend vermogen	2,46	2,43	2,21	2,30

De intraclasscorrelaties voor deze verschillende variabelen van het beleidsvoerend vermogen, laten zien dat de overtuigingen van leerkrachten gedeeld worden op schoolniveau. De respectievelijke scores voor gezamenlijke doelgerichtheid (ICC = 0,71), doeltreffende communicatie (ICC = 0,69) en reflecterend vermogen (ICC = 0,67) overschrijden telkens de drempelwaarde van 0,60. Dit laat toe om de opvattingen van leerkrachten met betrekking tot het beleidsvoerend vermogen te aggregeren op schoolniveau door het gemiddelde te nemen en dit als kenmerk van de school te beschouwen (Glick, 1985).

3. Beschrijving output-variabelen

De outputvariabelen verwijzen naar de evaluatiepraktijk van de bevraagde leerkrachten. Via het survey-onderzoek werd gepeild naar de mate waarin hun evaluatiepraktijk aanleunt bij het gescheiden, semi-geïntegreerd en het geïntegreerde model. De respondenten konden bij elk van deze modellen weergeven in welke mate dit representatief was voor hun praktijk door middel van een negenpuntenschaal waarbij nul staat voor ‘nooit’ en acht staat voor ‘altijd’.

Het **gescheiden model**, waarbij de instructieactiviteiten sterk gescheiden zijn van de evaluatieactiviteiten, komt het frequentste voor. Zowel de leerkrachten van het basis- als van het secundair onderwijs geven met gemiddelde scores van respectievelijk 5,11 en 5,51 aan dat dit model eerder vaak van toepassing is voor hun praktijk. Het **semi-geïntegreerde model** komt in mindere mate voor aangezien de respondenten dit model als soms representatief beschouwen voor hun praktijk. Dit geldt voor zowel het basis- als het secundair onderwijs, al is het zo dat het basisonderwijs (gemiddeld 3,81) hier iets hoger scoort dan het secundair onderwijs (gemiddeld 3,67). Het **geïntegreerde model**, ten slotte, is in vergelijking met de overige twee modellen het minste herkenbaar in de evaluatiepraktijk van de respondenten: leerkrachten geven aan dat dit model soms tot eerder weinig van toepassing is op hun praktijk. In het basisonderwijs (gemiddeld 3,04) komt dit model nog iets meer voor dan in het secundair onderwijs (gemiddeld 2,70).

De evaluatiepraktijk van leerkrachten is over het algemeen beperkt ingebed in de ruimere onderwijspraktijk van de leerkracht. In het basisonderwijs zijn er, in vergelijking met het secundair onderwijs, meer praktijken die aanleunen bij (semi-)geïntegreerde model.

Op schoolniveau zien we dat de evaluatiepraktijken van leerkrachten weinig gedeeld zijn. De intraclasscorrelaties zijn telkens lager dan 0,60. Concreet gaat het om de waarden 0,48 (gescheiden model) , -0,03 (semi-geïntegreerd model) en 0,18 (geïntegreerd model).

Hoofdstuk V Multivariate analyses

1. Werkwijze

Via de multivariate analyses willen we enerzijds meer inzicht verwerven in de drie modellen voor evaluatiepraktijk. Anderzijds willen we achterhalen welke precies de bepalende factoren zijn op zowel leerkracht- als schoolniveau voor de evaluatiepraktijken van leerkrachten. Hiertoe maken we respectievelijk gebruik van meervoudige lineaire regressies en multilevelanalyses. Voor deze analyses wordt verder gewerkt met de variabelen die eerder beschreven werden bij de univariate analyses, al is het zo dat er hierbij bepaalde keuzes gemaakt werden op basis van inhoudelijke, methodologische en/of pragmatische overwegingen. De onderstaande tabel biedt een overzicht van deze variabelen waarbij telkens weergegeven wordt (1) in welke multivariate analyse ze wel of niet meegenomen werden en (2) op basis van welke redenen.

VARIABLE	MULTIVARIATE ANALYSE		REDEN OF ARGUMENT
	Regressie	Multilevel	
LEERKRACHTNIVEAU			
Geslacht		X	Achtergrondvariabele
Leeftijd			Niet opgenomen omdat dit onderling sterk correleert met de variabele 'ervaring'.
Ervaring		X	De ervaring van leerkrachten nemen we op als onafhankelijke variabele in de multilevelanalyses om na te gaan in welke mate dit een bepalende factor is voor de evaluatiepraktijk van leerkrachten (zie onderzoeksvragen p. 24). Op basis van literatuur (p. 16) verwachten we meer ervaren leerkrachten beter in staat zijn om evaluatie te integreren in het ruimere onderwijsleerproces en bijgevolg sterker zullen aanleunen bij het geïntegreerde model. De bevindingen uit het casestudieonderzoek (Ysenbaert et al., 2018) wezen in die richting.
Vooropleiding		X	De vooropleiding van leerkrachten nemen we op als onafhankelijke variabele in de multilevelanalyses om na te gaan in welke mate dit een bepalende factor is voor de evaluatiepraktijk van leerkrachten (zie onderzoeksvragen p. 24). Literatuur (p. 15) stelt dat de vooropleiding van leerkrachten een rol speelt in hun praktijken.
Leerjaar/leeftijdsgroep			Op basis van theorie verwachten we hier geen effecten. Bovendien omvat deze categorische variabele veel categorieën (>6 verspreid over basis- en secundair onderwijs). Door deze variabele weg te laten houden we het statistisch model spaarzaam.
Vak			Op basis van theorie verwachten we hier geen effecten. Bovendien omvat deze categorische variabele veel categorieën (>6 verspreid over basis- en secundair onderwijs). Door deze variabele weg te laten houden we het statistisch model spaarzaam.
Andere functie			Op basis van theorie verwachten we hier geen effecten. Bovendien omvat deze categorische variabele veel categorieën (>6 verspreid over basis- en secundair onderwijs). Door deze variabele weg te laten houden we het statistisch model spaarzaam.
Differentiatiepraktijk	X	X	Op basis van literatuur (p. 12-13) verwachten we een positief verband tussen de algemene differentiatiepraktijk van leerkrachten en de mate waarin hun evaluatiepraktijk gekenmerkt wordt door het geïntegreerde model. Door deze variabele op te nemen als onafhankelijke variabele willen we de drie theoretisch ontwikkelde modellen voor evaluatiepraktijk (Ysenbaert et al., 2017) via multivariate regressieanalyse verder verfijnen en zo een antwoord formuleren op de onderzoeksvragen zoals beschreven op p.21 . Daarnaast nemen we deze variabele

			ook mee op in de multilevelanalyses om te zien hoe dit al dan niet samenhangt met andere variabelen in het model.
Professionalisering		X	De mate waarin de leerkracht inzet op professionalisering inzake evaluatie nemen we op als onafhankelijke variabele in de multilevelanalyses om na te gaan in welke mate dit een bepalende factor is voor de evaluatiepraktijk van leerkrachten (zie onderzoeksvragen p. 24). Op basis van literatuur (p.15) verwachten we dat leerkrachten die meer inzetten op professionalisering beter in staat zullen zijn om te evalueren volgens het geïntegreerde model.
Opvatting diversiteit		X	De opvattingen van leerkrachten inzake diversiteit nemen we op als onafhankelijke variabele in de multilevelanalyses om na te gaan in welke mate dit een bepalende factor is voor de evaluatiepraktijk van leerkrachten (zie onderzoeksvragen p. 24). Op basis van literatuur (p.15) verwachten we dat leerkrachten die overtuigd zijn van de rol die onderwijs heeft met betrekking tot diversiteit, eerder zullen evalueren volgens het geïntegreerde model.
Oog voor diversiteit		X	De bekwaamheid van leerkrachten om diversiteit onder leerlingen waar te nemen, nemen we op als onafhankelijke variabele in de multilevelanalyses om na te gaan in welke mate dit een bepalende factor is voor de evaluatiepraktijk van leerkrachten (zie onderzoeksvragen p. 24). Op basis van literatuur (p.15) verwachten we dat leerkrachten die in staat zijn om diversiteit onder leerlingen op te merken, meer zullen evalueren volgens het geïntegreerde model.
Doelmatigheidsbeleving		X	De doelmatigheidsbeleving van leerkrachten nemen we op als onafhankelijke variabele in de multilevelanalyses om na te gaan in welke mate dit een bepalende factor is voor de evaluatiepraktijk van leerkrachten (zie onderzoeksvragen p. 24). Op basis van literatuur (p.15) verwachten we dat leerkrachten met een hogere doelmatigheidsbeleving meer zullen evalueren volgens het geïntegreerde model.
Evaluatiepraktijk			
Gescheiden	X	X	Deze vormen telkens de outputvariabelen.
Semi-geïntegreerd	X	X	
Geïntegreerd	X	X	
Bij evaluatie aandacht voor:			
Impact	X		

Betrokkenheid	X		Op basis van literatuur (p. 12-13) achten we deze aspecten relevant omdat ze rechtstreeks verband houden met de leerling zelf en hierdoor ook van belang zijn voor een evaluatiepraktijk die rekening houdt met de diversiteit onder leerlingen. Door deze variabelen op te nemen als onafhankelijke variabelen willen we de drie theoretisch ontwikkelde modellen voor evaluatiepraktijk (Ysenbaert et al., 2017) via multivariate regressieanalyse verder verfijnen en zo een antwoord formuleren op de onderzoeksvragen zoals beschreven op p.21 .
Eerlijkheid	X		
Transparantie	X		
Verantwoording	X		
Opvattingen onderwijs			
Leerlinggericht		X	De onderwijsopvattingen van leerkrachten nemen we op als onafhankelijke variabelen in de multilevelanalyses om na te gaan in welke mate dit een bepalende factor is voor de evaluatiepraktijk van leerkrachten (zie onderzoeksvragen p. 24). Op basis van literatuur (p. 15) verwachten we dat leerlingegegerichte opvattingen positief samenhangen met evaluatiepraktijken die gekenmerkt worden door het geïntegreerde model.
Leerstofgericht		X	
Opvattingen evaluatie			
Informereren		X	De opvattingen van leerkrachten met betrekking tot evaluatie nemen we op als onafhankelijke variabele in de multilevelanalyses om na te gaan in welke mate dit een bepalende factor is voor hun evaluatiepraktijk (zie onderzoeksvragen p. 24).
Kritisch		X	
Geen impact op instructie		X	Op basis van literatuur (p. 14) verwachten we dat leerkrachten met een kritische houding ten aanzien van evaluatie meer zullen aanleunen bij het geïntegreerde model.
SCHOOLNIVEAU			
Ligging		X	Achtergrondvariabele
Onderwijsnet		X	Achtergrondvariabele
Onderwijsniveau		X	Het onderwijsniveau van de school nemen we op als onafhankelijke variabele in de multilevelanalyses om na te gaan in welke mate dit een bepalende factor is voor de evaluatiepraktijk van leerkrachten (zie onderzoeksvragen p. 24). Op basis van literatuur (p.18) verwachten we dat de evaluatiepraktijk van leerkrachten in het secundair onderwijs sterker gescheiden zal zijn. In het lager onderwijs verwachten we in vergelijking met het secundair onderwijs meer praktijken die aanleunen bij het (semi) -geïntegreerde model. De bevindingen van het casestudieonderzoek (Ysenbaert et al, 2018) dit.

Onderwijsvorm		X	<p>Het studieaanbod van de school nemen we op als onafhankelijke variabelen in de multilevelanalyses om na te gaan in welke mate dit een bepalende factor is voor de evaluatiepraktijk van leerkrachten (zie onderzoeksvragen p. 24).</p> <p>Op basis van literatuur (p. 18) verwachten we verschillen tussen evaluatiepraktijken in bijvoorbeeld aso- of bso-georiënteerde scholen. Zo verwachten we een positieve samenhang tussen bso en een geïntegreerde evaluatiepraktijk omdat deze onderwijsvorm nog sterker gericht is op vaardigheden.</p>
Leerlingencompositie		X	<p>Het studieaanbod van de school nemen we op als onafhankelijke variabelen in de multilevelanalyses om na te gaan in welke mate dit een bepalende factor is voor de evaluatiepraktijk van leerkrachten (zie onderzoeksvragen p. 24).</p> <p>Op basis van literatuur (p. 19) verwachten we verschillen tussen evaluatiepraktijken van scholen naargelang het leerlingenpubliek gekenmerkt wordt door diversiteit. Zo verwachten we dat een sterkere mate van diversiteit samenhangt met een sterkere mate van een gescheiden evaluatiemodel.</p>
Professionalisering		X	<p>De mate waarin de school inzet op professionalisering inzake evaluatie nemen we op als onafhankelijke variabele in de multilevelanalyses om na te gaan in welke mate dit een bepalende factor is voor de evaluatiepraktijk van leerkrachten (zie onderzoeksvragen p. 24).</p> <p>We verwachten dat leerkrachten van scholen die professionalisering bij hun leerkrachten aanmoedigen beter in staat zullen zijn om te evalueren volgens het geïntegreerde model.</p>
Schoolorganisatie			
Jaarklassensysteem		X	<p>De mate waarin de school haar leerlingen groepeert volgens het jaarklassensysteem nemen we op als onafhankelijke variabele in de multilevelanalyses om na te gaan in welke mate dit een bepalende factor is voor de evaluatiepraktijk van leerkrachten (zie onderzoeksvragen p. 24).</p> <p>Op basis van het casestudieonderzoek (Ysenbaert et al., 2018) verwachten we dat leerkrachten van scholen die minder strikt vasthouden aan het jaarklassensysteem meer zullen aanleunen bij het geïntegreerde model van evaluatie.</p>
(Her)examens			<p>De univariate analyses laten zien dat alle scholen voor secundair onderwijs gebruik maken van examens. Er is bijgevolg geen variatie, waardoor het weinig zinvol is om dit mee op te nemen.</p>
Frequentie klassenraden			

Frequentie oudercontacten			<p>De univariate analyses laten zien dat alle bevraagde scholen minstens twee keer per schooljaar klassenraden en oudercontacten organiseren. De variatie in frequentie is vrij beperkt. Dit in tegenstelling tot de organisatie van leerlingcontacten dat varieert van nooit tot meer dan vijf keer per schooljaar.</p> <p>Rekening houdend met de spaarzaamheid van het multilevelmodel nemen we enkel de frequentie van de leerlingcontacten mee in de multilevelanalyses.</p> <p>Bovendien verwachten we een positieve samenhang tussen leerlingcontact en een geïntegreerde evaluatiepraktijk aangezien het geïntegreerde model aansluit bij een meer constructivistische opvatting op leren die de leerling meer actief betreft (zie theoretisch kader p 11-12.) .</p>
Frequentie leerlingcontacten		X	
Gebruik leerlingvolgsysteem			<p>Alle scholen maken gebruik van een leerlingvolgsysteem. Er is bijgevolg geen variatie op vlak van gebruik, waardoor het weinig zinvol is om dit mee op te nemen.</p> <p>Verspreid over basis- en secundair onderwijs omvat deze categorische variabele veel categorieën (>6). In functie van de spaarzaamheid van het model nemen we deze variabele niet op bij de multivariate analyses.</p>
Gebruik van het rapport			In functie van de spaarzaamheid van het model nemen we het aspect van 'rapport' (dat in werkelijkheid meerdere deelvariabelen omvat) niet op.
Gebruik gevalideerde toetsen			De univariate analyses laten zien dat alle scholen onderwijs gebruik maken van examens. Er is bijgevolg geen variatie, waardoor het weinig zinvol is om dit mee op te nemen.
Beleidsvoerend vermogen:			<p>Met betrekking tot 'beleidsvoerend vermogen' selecteerden we voor de spaarzaamheid van het model slechts één van de drie variabelen (die overigens onderling sterk correleren). We kozen hier voor 'gezamenlijke doelgerichtheid' omdat dit de hoogste intraclasscorrelatie heeft (ICC= 0.71).</p>
Gezamenlijke doelgerichtheid		X	
Doeltreffende communicatie			
Reflecterend vermogen			

2. Modellen voor evaluatiepraktijk

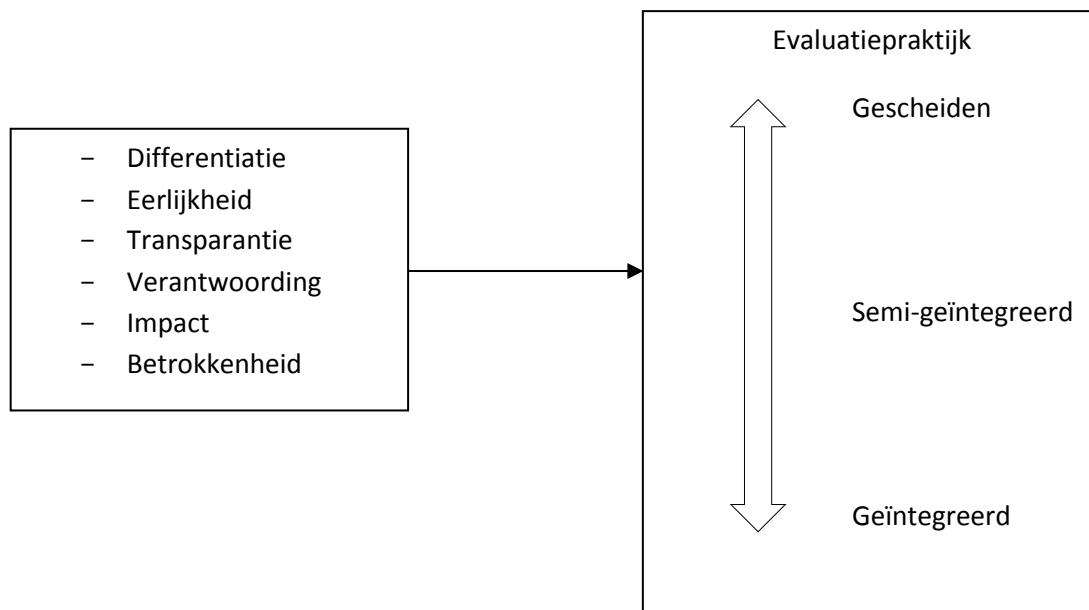


Fig. 6: Conceptueel model voor evaluatiepraktijk

Op basis van het theoretisch kader kwamen we tot bovenstaand conceptueel model waarbij we volgende onderzoeksvragen formuleerden:

- Hoe hangt een gescheiden evaluatiepraktijk samen met een sterkere differentiatiepraktijk in het algemeen en met de verschillende kwaliteitsaspecten voor evaluatie zoals 'eerlijkheid', 'transparantie', 'verantwoording', 'impact' en 'betrokkenheid'?
- Hoe hangt een geïntegreerde evaluatiepraktijk samen met een sterkere differentiatiepraktijk in het algemeen en met de verschillende kwaliteitsaspecten voor evaluatie zoals 'eerlijkheid', 'transparantie', 'verantwoording', 'impact' en 'betrokkenheid'?
- Hoe hangt een semi-geïntegreerde evaluatiepraktijk samen met een sterkere differentiatiepraktijk in het algemeen en met de verschillende kwaliteitsaspecten voor evaluatie zoals 'eerlijkheid', 'transparantie', 'verantwoording', 'impact' en 'betrokkenheid'?
- Kunnen de drie modellen voor evaluatiepraktijk geplaatst worden op een continuüm waarbij de effecten voor de semi-geïntegreerde evaluatiepraktijk minder uitgesproken zijn dan bij de gescheiden of geïntegreerde evaluatiepraktijk?
- Is er sprake van een schoolcultuur met betrekking tot de wijze waarop er geëvalueerd wordt? Zijn de evaluatiepraktijken van leerkrachten op eenzelfde school dermate gedeeld dat we kunnen stellen dat de evaluatiepraktijk op een school gekenmerkt wordt door één van de drie modellen?

Uit de beschrijvende analyses weten we reeds dat er geen sprake is van een schoolcultuur met betrekking tot de wijze waarop er geëvalueerd wordt. De evaluatiepraktijken van leerkrachten kunnen niet als een schoolkenmerk gezien worden omdat de evaluatiepraktijken van leerkrachten onderling veel verschillen.

Om een antwoord te formuleren op de overige onderzoeksvragen maken we gebruik van meervoudige lineaire regressies waarbij we de verschillende modellen voor evaluatiepraktijk telkens opnemen als afhankelijke variabele. Als onafhankelijke variabelen nemen we de verschillende kwaliteitsaspecten met betrekking tot evaluatie mee alsook de mate waarin de leerkracht inzet op differentiatie in de algemene klaspraktijk.

2.1. Gescheiden model als afhankelijke variabele

In tabel 5.1 zien we dat een gescheiden of niet-ingebedde evaluatiepraktijk samenhangt met aandacht voor de kwaliteitsaspecten ‘verantwoording’ en ‘betrokkenheid’. Het verband met een evaluatiepraktijk die aandacht heeft voor het verantwoordingsaspect is matig positief. Het verband met aandacht voor betrokkenheid is eerder negatief. Dit laatste verband is eerder zwak. We stellen verder een significant negatief verband vast met de algemene differentiatiepraktijk van leerkrachten. Een gescheiden evaluatiepraktijk wordt dus voorspeld door een sterkere gerichtheid op verantwoording, wat inhoudt dat de leerkracht meer aandacht besteedt aan het kunnen toelichten waar een bepaald resultaat op een evaluatie vandaan komt. Er blijkt verder een beperktere aandacht te zijn voor de betrokkenheid van leerlingen bij het evaluatieproces en voor differentiatie in het algemeen.

Er is geen significant verband met de variabelen ‘eerlijkheid’, ‘transparantie’ en ‘impact’. De coëfficiënten van deze variabelen wijzen telkens op een eerder positieve samenhang met het model voor gescheiden evaluatiepraktijk.

Tabel 5.1

Resultaten van de meervoudige regressieanalyse op het model voor gescheiden evaluatiepraktijk

Gescheiden evaluatiepraktijk			
Variabele	B	SE	β
Intercept	3,828	,480	
Aandacht voor kwaliteitscriteria:			
Eerlijkheid	,207	,117	,074
Transparantie	,060	,124	,022
Verantwoording	,576	,153	,179***
Impact	,006	,131	,002
Betrokkenheid	-,225	,092	-,113*
Differentiatiepraktijk	-,224	,064	-,153***

Noot: * $p < ,05$, ** $p < ,01$, *** $p < ,001$

2.2. Geïntegreerd model als afhankelijkke

Met betrekking tot de geïntegreerde evaluatiepraktijk zien we in tabel 5.2 een verband met de ‘aandacht voor betrokkenheid van leerlingen’ en ‘differentiatie’. Beide verbanden zijn positief en zijn respectievelijk zwak tot gematigd van aard. Dit betekent dat hoe meer een leerkracht inzet op differentiatie in het algemeen en op het betrekken van leerlingen tijdens de evaluatie, de evaluatiepraktijk vaker gekenmerkt wordt door een sterke integratie in het onderwijsleerproces.

Voor de overige aspecten ‘eerlijkheid’, ‘transparantie’, ‘verantwoording’ en ‘impact’ werd geen significant verband vastgesteld. De coëfficiënten laten zien dat er eerder een negatieve samenhang is met deze variabelen.

Tabel 5.2

Resultaten van de meervoudige regressieanalyse op het model voor geïntegreerde evaluatiepraktijk

Geïntegreerde evaluatiepraktijk			
Variabele	B	SE	β
Intercept	1,162	,639	
Aandacht voor kwaliteitscriteria:			
Eerlijkheid	-,070	,155	-,018
Transparantie	-,215	,165	-,056
Verantwoording	-,038	,203	-,009
Impact	-,014	,175	-,004
Betrokkenheid	,485	,123	,177***
Differentiatiepraktijk	,549	,085	,273***

Noot: * $p < ,05$, ** $p < ,01$, *** $p < ,001$

2.3. Semi-geïntegreerd model als afhankelijke

Bij een semi-geïntegreerde evaluatiepraktijk blijken ‘differentiatiepraktijk’, ‘aandacht voor impact’ en ‘aandacht voor betrokkenheid’ voorspellende variabelen te zijn (tabel 5.3). Het verband is voor elke van deze drie variabelen eerder zwak en positief. Dit wijst erop dat leerkrachten wiens evaluatiepraktijk deels ingebed is in de ruimere onderwijspraktijk meer aandacht hebben voor de impact van evaluatie op het onderwijsleerproces en de betrokkenheid van leerlingen tijdens het

evaluatieproces. Bovendien verwachten we dat de evaluatiepraktijk van leerkrachten die meer inzetten op differentiatie in het algemeen, ook meer inhaakt op de onderwijsactiviteiten.

Er werd geen significant verband vastgesteld met de variabelen ‘eerlijkheid’, ‘transparantie’ en ‘verantwoording’. De coëfficiënt voor ‘transparantie’ wijst, net zoals bij het ‘gescheiden model’, op een eerder negatieve samenhang, maar minder sterk in vergelijking met het geïntegreerde model. De coëfficiënt voor ‘eerlijkheid’ wijst eveneens op een eerder negatieve samenhang, die sterker is dan bij het geïntegreerde model. Uit de coëfficiënt voor ‘verantwoording’ leiden we af dat de samenhang met een semi-geïntegreerde praktijk licht positief is.

Tabel 5.3

Resultaten van de meervoudige regressieanalyse op het model voor semi-geïntegreerde evaluatiepraktijk

Variabele	Semi-geïntegreerde evaluatiepraktijk		
	B	SE	β
Intercept	1,633	,580	
Aandacht voor kwaliteitscriteria			
Eerlijkheid	-,219	,141	-,064
Transparantie	-,020	,150	-,006
Verantwoording	,008	,185	,002
Impact	,311	,159	,089*
Betrokkenheid	,409	,111	,168***
Differentiatiepraktijk	,343	,077	,191***

*Noot: *p < ,05, **p < ,01, ***p < ,001*

2.4. Conclusie met betrekking tot modellen voor evaluatiepraktijk

Een gescheiden evaluatiepraktijk hangt negatief samen met de algemene differentiatiepraktijk en de mate waarin leerlingen betrokken worden bij het evaluatieproces. Bij leerkrachten die sterk inzetten op een gescheiden evaluatiepraktijk zullen we weinig differentiatiepraktijken aantreffen in hun lessen en zullen de leerlingen ook niet betrokken worden bij het proces van evaluatie. Opmerkelijk is dat er een significant positief verband is voor het aspect ‘verantwoording’. Dit betekent dat leerkrachten wiens evaluatiepraktijk gekenmerkt wordt door een strikte scheiding met de instructie meer aandacht besteden aan het verantwoorden aan verschillende actoren waar een bepaald resultaat van een evaluatie vandaan komt. Deze positieve samenhang zou verklaard kunnen worden door de eenvoud van dit model wat het eventueel ‘makkelijker’ maakt om erover te communiceren met andere actoren.

Het geïntegreerde model daarentegen hangt positief samen met een algemene evaluatiepraktijk en het betrekken van leerlingen bij evaluatie, wat lijnrecht staat tegenover de bevindingen bij het gescheiden model. Hoe meer een leerkracht tijdens zijn/haar lespraktijk aandacht heeft voor de verschillen tussen leerlingen en hoe meer leerlingen betrokken worden bij het evaluatieproces, des te sterker de evaluatiepraktijk ingebed zal zijn in de instructie.

Het semi-geïntegreerde model toont eveneens een positieve samenhang met de algemene differentiatiepraktijk en de betrokkenheid van leerlingen bij evaluatie. Bovendien is er ook een positieve samenhang met de aandacht voor ‘impact’. Dit laatste kan verklaard worden door het feit

dat de impact van evaluatie op de instructie expliciet tot uiting komt in de representatie van het semi-geïntegreerde model.

Er is dus een duidelijk continuüm aanwezig voor de aspecten ‘betrokkenheid’ en ‘differentiatiepraktijk’ die telkens significante verbanden opleverden, en dit in positieve richting voor de geïntegreerde evaluatiepraktijk en in negatieve richting voor de gescheiden evaluatiepraktijk. Een geïntegreerde evaluatiepraktijk komt veel meer voor in combinatie met een klaspraktijk die inzet op differentiatie waarbij leerlingen sterker betrokken worden bij het evaluatieproces. Deze aspecten zijn significant minder aanwezig bij een semi-geïntegreerde evaluatiepraktijk en nog minder bij een gescheiden evaluatiepraktijk.

Het is opmerkelijk dat er bij geen enkel model significanties gevonden worden voor ‘eerlijkheid’ en ‘transparantie’. Uit de univariate analyses stellen we echter vast dat de leerkrachten hier wel aandacht aan besteden. De gemiddelde scores op een schaal van nul tot vier voor basis- en secundair onderwijs zijn respectievelijk 2,70 en 3,90 voor ‘eerlijkheid’ en 3,09 en 3,42 voor ‘transparantie’. We kunnen dus besluiten dat het hebben van aandacht voor aspecten zoals eerlijkheid en transparantie niet zozeer samenhangt met (één van) deze modellen en dat leerkrachten hier sowieso aandacht voor hebben ongeacht het model waardoor hun evaluatiepraktijk gekenmerkt wordt.

3. Multilevelanalyses

3.1. Methode

Voor verdere verklarende analyses gaan we na welke de verklarende factoren zijn voor de evaluatiepraktijken van leerkrachten. Hiertoe wordt gebruik gemaakt van multilevelanalyses omdat deze rekening houden met het feit dat er meerdere niveaus zijn. Leerkrachten (niveau 1) maken immers steeds deel uit van scholen (niveau 2). De antwoorden van de leerkrachten op de vragenlijsten worden niet als louter onafhankelijk gezien. Er wordt rekening mee gehouden dat de antwoorden van leerkrachten uit eenzelfde school mogelijk meer overeenkomsten tonen dan bij leerkrachten van verschillende scholen. We maken hiervoor gebruik van het statistisch programma MLwiN 3.04.

3.1.1. Afhankelijke variabelen

We zijn voornamelijk geïnteresseerd in de evaluatiepraktijk van leerkrachten en dit zal dan ook de afhankelijke variabele vormen bij de analyses. Deze evaluatiepraktijk hebben we geoperationaliseerd aan de hand van drie variabelen die telkens verwijzen naar (1) het model van een gescheiden evaluatiepraktijk die gekenmerkt wordt door een sterkere gerichtheid op verantwoording, (2) een semi-geïntegreerde evaluatiepraktijk en (3) een geïntegreerde evaluatiepraktijk die gekenmerkt wordt door een sterkere betrokkenheid van leerlingen bij het evaluatieproces en meer aandacht voor differentiatie in de algemene klaspraktijk.

Elk van deze drie variabelen – gescheiden, semi-geïntegreerd en geïntegreerd - zal fungeren als afhankelijke variabele bij de multilevelanalyses.

3.1.2. Onafhankelijke variabelen

Voor de onafhankelijke variabelen maken we een onderscheid tussen de variabelen op het niveau van de leerkracht en op het niveau van de school. Hieronder worden de variabelen met betrekking tot het leerkrachtniveau toegelicht door te specificeren hoe we ze operationaliseren en met welke afkorting we er verder zullen naar verwijzen bij het toelichten van de resultaten. Wat de categorische variabelen betreft, wordt de referentiecategorie aangeduid met *.

- Geslacht: dichotoom, man*/vrouw
- Ervaring : aantal jaar ervaring in het onderwijs
- Opleiding: dichotoom, niet*/wel in het bezit van een masterdiploma of vergelijkbaar equivalent
- Onderwijsvorm: dichotoom, niet*/wel een lesopdracht in het bso
- Professionalisering: professionalisering met betrekking tot de evaluatiepraktijk (zie schaal zoals behandeld in 2.3)
- Doelmatigheidsbeleving: het geloof in het eigen kunnen als leerkracht (zie *self-efficacy*-schaal zoals behandeld in 2.8)
- Bekwaamheidsgevoel met betrekking tot het waarnemen van diversiteit: de mate waarin men oog heeft voor diversiteit (zie schaal zoals behandeld in 2.6)
- Onderwijsopvatting leerlinggericht: de mate waarin men er leerlinggerichte onderwijsopvattingen op na houdt (zie schaal zoals behandeld in 2.4)
- Onderwijsopvatting leerstofgericht: de mate waarin men er leerstofgerichte onderwijsopvattingen op na houdt (zie schaal zoals behandeld in 2.4)
- Opvattingen diversiteit: de mate waarin men ervan overtuigd is dat omgaan met diversiteit deel uitmaakt van zijn/haar kerntaak (zie schaal zoals behandeld in 2.5)
- Opvattingen evaluatie informeren: de mate waarin men ervan overtuigd is dat evaluatie een informerende functie heeft (zie schaal zoals behandeld in 2.7)
- Opvattingen evaluatie kritisch: de mate waarin men zich kritisch opstelt ten aanzien van evaluatie (zie schaal zoals behandeld in 2.7)
- Opvattingen evaluatie geen impact: de mate waarin men ervan overtuigd is dat evaluatie geen impact heeft op de onderwijspraktijk (zie schaal zoals behandeld in 2.7)

De variabelen op schoolniveau omvatten:

- Ligging: dichotoom, school ligt niet*/wel in een centrumstad
- Onderwijsniveau: dichotoom, basisonderwijs*/secundair onderwijs
- Onderwijsnet: vrij onderwijs*/gesubsidieerd onderwijs/officieel gesubsidieerd onderwijs
- Leerlingencompositie: het aandeel leerlingen met hoge SES (bepaald door aantal leerlingen op school die niet aantikken op de indicator 'laag opleidingsniveau moeder' of 'studietoelage')
- Studieaanbod: dichotoom, zonder*/mét een studieaanbod in de onderwijsvorm bso
- Beleidsvoerend vermogen van de school inzake 'evaluatie': de mate waarin er sprake is van gezamenlijke doelgerichtheid op school als het gaat over evalueren van leerlingen (zie schaal zoals behandeld in 2.10)
- Frequentie waarmee er op school leerlingencontacten georganiseerd worden: dichotoom, minder dan twee keer per schooljaar/twee of meerdere keren per schooljaar*

- Frequentie waarmee er klassenraden georganiseerd worden: dichotoom, klassenraden worden minder dan vijf keer per schooljaar georganiseerd/ klassenraden worden vijf of meerdere keren per schooljaar georganiseerd*
- Groepering: dichotoom, de leerlingen worden gegroepeerd volgens een strikte invulling van het jaarklassensysteem*/soms wordt er afgeweken van het jaarklassensysteem voor het groeperen van leerlingen
- Professionaliseringsbeleid: de mate waarin de school gekenmerkt wordt door een professionaliseringsbeleid dat specifiek inzet op het professionaliseren van de leerkrachten omtrent de evaluatie van leerlingen (zie aggregaten van schaal zoals behandeld in 2.3).

3.2. Modelopbouw

Voor het opbouwen van het statistisch model hanteren we de logica van het CIPO-model waarbij we na het nulmodel, achtereenvolgens de context-, input- en procesvariabelen zullen toevoegen. Hierbij nemen we telkens eerst de variabelen op schoolniveau mee en vervolgens de leerkrachtvariabelen.

In een eerste model voegen we de contextvariabelen toe, ofwel de onveranderlijke kenmerken op het niveau van de school: de ligging, het onderwijsniveau en het onderwijsnet. In een tweede model voegen we de inputvariabelen toe op het niveau van de school: de leerlingcompositie en het studieaanbod. De inputvariabelen op het niveau van de leerkracht worden toegevoegd in een derde model: geslacht, ervaring en vooropleiding. De toevoeging van de procesvariabelen worden gespreid over meerdere modellen. In het vierde model worden de procesvariabelen op schoolniveau toegevoegd: de wijze van organisatie, de frequentie van leerlingcontacten en klassenraden en het beleidsvoerend vermogen. Gezien het grote aantal aan procesvariabelen op leerkrachtniveau worden deze er in twee stappen aan toegevoegd. In een eerste stap (model 5) worden de variabelen die verwijzen naar het inschatten van het bekwaamheidsgevoel opgenomen: doelmatigheidsbeleving, bekwaamheidsgevoel met betrekking tot het oog hebben voor diversiteit en de mate waarin men zich professionaliseert inzake evaluatie. In een laatste stap (model 6) worden de variabelen die verwijzen naar de opvattingen van leerkrachten toegevoegd: de leerlinggerichte en leerstofgerichte onderwijsopvattingen, de opvattingen omtrent diversiteit, de opvattingen omtrent evaluatie (informereren, kritisch en geen impact).

3.3. Resultaten

3.3.1. Gescheiden evaluatiepraktijk als afhankelijke variabele

Model 0: het nulmodel

Het nulmodel omvat een model met twee niveaus en een random intercept met 'gescheiden evaluatiepraktijk' als uitkomstvariabele. Het intercept representeert de gemiddelde score voor 'gescheiden evaluatiepraktijk' voor alle leerkrachten die de vragenlijst invulden uit alle scholen die deelnamen aan het onderzoek. Uit dit model blijkt dat de gemiddelde score 5,34 (SD = 0.08) eerder hoog is wat erop wijst dat de evaluatiepraktijk van leerkrachten eerder vaak gekenmerkt wordt door een strikte scheiding tussen evaluatie en instructie.

De *variance partitioning coefficient* (VPC) heeft een waarde van 0.063. Dit toont aan dat 6,3% van de totale resterende variantie zich op schoolniveau bevindt.

Model 1: contextvariabelen

In model 1 werden de schoolvariabelen ligging, onderwijsnet en onderwijsniveau toegevoegd als extra verklarende variabelen. Een basisschool behorend tot het vrije onderwijsnet en niet in een centrumstad ligt wordt gezien als referentiecategorie. Ligging en onderwijsnet hebben geen significant effect op het meer of minder hanteren van een gescheiden evaluatiepraktijk door leerkrachten. Het onderwijsniveau blijkt wel een positief significant effect te hebben ($p = 0.03$). De mate waarin leerkrachten een gescheiden evaluatiepraktijk hanteren zal gemiddeld genomen 0.33 punten toenemen voor leerkrachten in het secundair onderwijs in vergelijking met leerkrachten in het basisonderwijs.

Model 2: inputvariabelen (schoolniveau)

Omwillen van de hoeveelheid aan variabelen worden de niet-significante contextvariabelen, onderwijsnet en ligging, niet opgenomen in model 2. De variabele onderwijsniveau blijft behouden en de inputvariabelen op schoolniveau, leerlingencompositie en studieaanbod, worden toegevoegd. We maken een onderscheid tussen model 2a en model 2b waarbij studieaanbod (referentiecategorie: zonder aanbod bso) toegevoegd wordt in een afzonderlijk model omdat deze variabele enkel van toepassing is op de scholen voor secundair onderwijs. Hiermee willen we rekening houden met het feit dat MIWin *listwise deletion* toepast, waardoor alle data die betrekking heeft op het basisonderwijs als *missing* behandeld wordt. Om een vertekend beeld te voorkomen, voegen we deze variabele toe in een afzonderlijk model 2b.

De schoolvariabelen leerlingencompositie en studieaanbod hebben geen significant effect op de mate waarin de evaluatiepraktijk van een leerkracht gekenmerkt wordt door een gescheiden evaluatiepraktijk. Het geringe significante effect van onderwijsniveau blijft behouden voor model 2a, in model 2b is dit niet meer relevant aangezien het om uitsluitend secundair onderwijs gaat.

Hoewel niet significant behouden we de variabele leerlingencompositie bij de volgende stappen omdat we niet alleen geïnteresseerd zijn in het effect van leerlingencompositie op de evaluatiepraktijk van leerkrachten, maar ook verwachten dat dit effect mogelijk kan wijzigen door eventuele samenhang met procesvariabelen.

Model 3: inputvariabelen (leerkrachtniveau)

Voor model 3a worden de inputvariabelen op leerkrachtniveau, geslacht, ervaring en vooropleiding toegevoegd aan model 2a. Geen van deze nieuwe variabelen heeft een significant effect op de evaluatiepraktijk van leerkrachten. We zien tevens dat het positief significante effect van onderwijsniveau wegvalt na toevoeging van de inputvariabelen op leerkrachtniveau.

In model 3b gaan we voor de scholen secundair onderwijs na of de onderwijsvorm waarin leerkrachten les geven een effect heeft op de evaluatiepraktijk. We stellen een significant effect vast voor het bso ($p = 0.01$): de evaluatiepraktijk van leerkrachten die les geven in het bso zal minder vaak aansluiten bij het gescheiden model dan leerkrachten die geen les geven in het bso.

Model 4: procesvariabelen (schoolniveau)

In model 4 wordt verder gewerkt op model 3a na het weglaten van de niet-significante inputvariabelen op leerkrachtniveau (geslacht, ervaring en vooropleiding). Aan de hand van het vierde model gaan we het effect van de procesvariabelen op schoolniveau na: het beleidsvoerend vermogen van de school inzake leerlingenevaluatie, de wijze waarop leerlingen gegroepeerd worden, en de frequentie van klassenraden en leerlingencontacten. We merken hierbij twee significante effecten op. Er is een positief effect voor het beleidsvoerend vermogen van de school inzake leerlingenevaluatie ($p = 0.04$): indien het beleidsvoerend vermogen van de school met één punt toeneemt dan neemt de mate waarin de evaluatiepraktijk van een leerkracht van die school aansluit bij het gescheiden model toe met een score van 0.54. We merken een tweede positief effect op voor de frequentie waarmee leerlingencontacten georganiseerd worden ($p = 0.03$). Leerkrachten van scholen die minstens twee keer op een schooljaar een leerlingencontact organiseren, leunen meer (met een score van 0.37) aan bij een gescheiden evaluatiepraktijk dan leerkrachten van scholen waar er minder dan twee leerlingencontacten per schooljaar zijn.

Het effect van onderwijsniveau is in model 4 opnieuw significant ($p = 0.01$) na het weglaten van de leerkrachtvariabelen. Dit wijst erop dat de verklarende kracht eerder bij de leerkrachten zelf gesitueerd is, die wellicht gemeenschappelijke kenmerken vertonen indien ze tot eenzelfde onderwijsniveau behoren, dan dat de verklaring bij het onderwijsniveau zelf gezocht moet worden.

Model 5: procesvariabelen (leerkrachtniveau)

In model 5 voegen we de procesvariabelen op het niveau van de leerkracht toe. We laten de niet-significante procesvariabelen op schoolniveau weg en nemen de leerkrachtvariabelen geslacht, ervaring en vooropleiding terug op. Aangezien het om heel wat procesvariabelen gaat, doen we dit stapsgewijs: model 5a focust op de variabelen die verwijzen naar de praktijk van leerkrachten, de overige variabelen die verwijzen naar de opvattingen van leerkrachten worden extra toegevoegd in model 5b.

In model 5a merken we geen significanties op voor de mate van professionalisering, de doelmatigheidsbeleving en het bekwaamheidsgevoel om oog te hebben voor diversiteit onder de leerlingen. De positieve significante effecten van onderwijsniveau en leerlingencontact blijven (beide $p = 0.02$). Het positieve effect van het beleidsvoerend vermogen verdwijnt na toevoegen van deze procesvariabelen.

In model 5b worden de opvattingen van de leerkrachten toegevoegd. We stellen niet-significante effecten vast voor de onderwijsopvattingen van leerkrachten (leerlinggericht en leerstofgericht) en de opvattingen met betrekking tot diversiteit. De opvattingen met betrekking tot evaluatie zijn eveneens

niet significant voor de mate waarin men ervan overtuigd is dat evaluatie geen impact heeft of men er kritisch tegenover staat. Er is een positief significant effect merkbaar voor de opvattingen van leerkrachten die verwijzen naar de mate waarin men overtuigd is van de informerende waarde van evaluatie ($p = 0.001$). Indien de overtuiging van een leerkracht over de informerende waarde van evaluatie met één punt toeneemt, dan neemt de score voor het aansluiten bij een gescheiden evaluatiepraktijk toe met een waarde van 0.45.

Het significante effect van onderwijsniveau wordt wegverklaard na toevoegen van de variabelen die verwijzen naar de opvattingen van leerkrachten. Het positieve effect van het leerlingencontact blijft behouden ($p = 0.04$) na het meenemen van de opvattingen van leerkrachten.

Tabel 5.4

Resultaten van de multilevelanalyses met 'gescheiden evaluatiepraktijk' als afhankelijke

	Model 0	Model 1	Model 2a	Model 2b	Model 3a
FIXED PART					
Intercept	5.34 (0.08)***	5.24 (0.14)***	5.16 (0.12)***	5.86 (0.20)***	5.16 (0.12)***
SCHOOL					
Ligging		-0.25(0.16)			
Net:					
GO		0.17 (0.18)			
OGO		-0.18 (0.31)			
Niveau		0.33 (0.15)*	0.34 (0.16)*	0.00(0.00)	0.28 (0.17)
Compositie			-0.13 (0.30)	-0.07 (0.33)	-0.10 (0.29)
Aanbod bso				-0.37 (0.21)	
Beleidsvoerend vermogen					
Leerlingencontact					
Klassenraad					
Groepering					
LEERKRACHT					
Geslacht					-0.01 (0.14)
Ervaring					-0.01 (0.00)
Opleiding					0.15 (0.14)
Professionalisering					
Doelmatigheid					
Oog voor diversiteit					
Onderwijsopv. leerling					
Onderwijsopv. leerstof					
Opvatting diversiteit					
Opv. evaluatie informeren					
Opv. evaluatie kritisch					
Opv. evaluatie gn impact					
RANDOM PART					
School	0.18 (0.07)	0.10 (0.06)	0.13 (0.06)	0.00 (0.00)	0.13 (0.06)
Leerkracht	2.62 (0.13)	2.63 (0.13)	2.63 (0.13)	2.23 (0.16)	2.62 (0.13)
N (scholen)	56	56	56	22	56
N (leerkrachten)	816	816	816	398	816

*Noot: Per cel wordt weergegeven: regressiecoëfficiënt (standaardfout); * $p < ,05$, ** $p < ,01$, *** $p < ,001$*

Tabel 5.4 (vervolg)

Resultaten van de multilevelanalyses met 'gescheiden evaluatiepraktijk' als afhankelijke

	Model 3b	Model 4	Model 5a	Model 5b
FIXED PART				
Intercept	5.72 (0.12)***	4.83 (0.19)***	4.79 (0.19)***	4.92 (0.17)***
SCHOOL				
Ligging				
Net:				
GO				
OGO				
Niveau	0.00(0.00)	0.63 (0.23)**	0.47 (0.20)*	0.35 (0.18)
Compositie	0.02 (0.27)	-0.29 (0.31)	-0.30 (0.32)	-0.18 (0.30)
Aanbod bso	-0.40 (0.14)**			
Beleidsvoerend vermogen		0.54 (0.26)*	0.50 (0.26)	0.43 (0.24)
Leerlingencontact		0.37 (0.18)*	0.40 (0.18)*	0.35 (0.16)*
Klassenraad		0.05 (0.20)		
Groepering		-0.17 (0.18)		
LEERKRACHT				
Geslacht	-0.18 (0.15)		-0.13 (0.16)	-0.12 (0.16)
Ervaring	-0.01 (0.01)		-0.01 (0.01)	-0.01 (0.01)
Opleiding	0.04 (0.14)		0.22 (0.15)	0.22 (0.15)
Professionalisering			0.10 (0.14)	0.03 (0.14)
Doelmatigheid			0.15 (0.10)	0.03 (0.10)
Oog voor diversiteit			-0.11 (0.06)	-0.09 (0.06)
Onderwijsopv. leerling				-0.19 (0.19)
Onderwijsopv. leerstof				0.23 (0.16)
Opvatting diversiteit				-0.14 (0.11)
Opv. evaluatie informeren				0.45 (0.14)**
Opv. evaluatie kritisch				-0.07 (0.10)
Opv. evaluatie gn impact				-0.03 (0.10)
RANDOM PART				
School	0.00 (0.00)	0.10 (0.06)	0.07 (0.05)	0.07 (0.05)
Leerkracht	2.27 (0.14)	2.58 (0.14)	2.38 (0.13)	2.37 (0.13)
N (scholen)	29	47	47	47
N (leerkrachten)	493	706	706	706

Noot: Per cel wordt weergegeven: regressiecoëfficiënt (standaardfout); * $p < ,05$, ** $p < ,01$, *** $p < ,001$

3.3.2. Semi-geïntegreerde evaluatiepraktijk als afhankelijke variabele

Model 0: het nulmodel

Het nulmodel omvat een model met twee niveaus en een random intercept met 'semi-geïntegreerde evaluatiepraktijk' als uitkomstvariabele. Het intercept representeert de gemiddelde score voor 'gescheiden evaluatiepraktijk' voor alle leerkrachten die de vragenlijst invulden uit alle scholen die deelnamen aan het onderzoek. Uit dit model blijkt dat de gemiddelde score 3.73 (SD = 0.07) is. Dit wijst erop dat de evaluatiepraktijk van leerkrachten soms gekenmerkt wordt door een evaluatiepraktijk die inhaakt op de activiteit van instructie.

De *variance partitioning coefficient* (VPC) heeft een waarde van 0.001 wat erop wijst dat slechts 0,1% van de totale resterende variantie zich op schoolniveau bevindt.

Model 1: contextvariabelen

De schoolvariabelen ligging, onderwijsnet en onderwijsniveau werden toegevoegd als bijkomende verklarende variabelen in model 1. Een basisschool behorend tot het vrije onderwijsnet en niet gelegen in een centrumstad wordt gezien als referentiecategorie. Ligging, onderwijsnet en onderwijsniveau hebben geen significant effect op het meer of minder hanteren van een semi-geïntegreerde evaluatiepraktijk door leerkrachten.

Model 2: inputvariabelen (schoolniveau)

Omwillen van de hoeveelheid aan variabelen worden de contextvariabelen, onderwijsnet en ligging, niet opgenomen in model 2. De variabele onderwijsniveau blijft voorlopig behouden en de inputvariabelen op schoolniveau, leerlingencompositie en studietoeraanbod, worden toegevoegd. Naar analogie met de modelopbouw waarbij de gescheiden evaluatiepraktijk de afhankelijke variabele vormde, maken we een onderscheid tussen model 2a en model 2b waarbij studietoeraanbod (referentiecategorie: zonder aanbod bso) toegevoegd wordt in een afzonderlijk model omdat deze variabele enkel van toepassing is op de scholen voor secundair onderwijs. Dit om te voorkomen dat we een vertekend beeld zouden krijgen omdat studietoeraanbod enkel van toepassing is op de scholen voor secundair onderwijs.

In model 2a stellen we een significant effect vast voor leerlingencompositie ($p = 0.005$). Dit effect is negatief: hoe groter het aandeel hoge SES leerlingen op een school, hoe minder de evaluatiepraktijk van leerkrachten gekenmerkt wordt door een semi-geïntegreerde evaluatiepraktijk. Indien het aandeel leerlingen met hoge SES met één punt stijgt, dan neemt de waarde voor semi-geïntegreerde evaluatiepraktijk met -0.70 af.

Model 2b geeft een niet-significant effect weer voor 'studietoeraanbod' ($p = 0.30$). Bovendien verdwijnt het effect van leerlingencompositie ($p = 0.065$) na het controleren voor studietoeraanbod.

In de volgende modellen behouden we de variabele 'leerlingencompositie' en nemen we 'studieaanbod' verder niet mee.

Model 3: inputvariabelen (leerkrachtniveau)

Voor model 3a bouwen we verder op model 2a, waarbij we de inputvariabelen op leerkrachtniveau, geslacht, ervaring en vooropleiding toevoegen. Het model toont geen significante effecten voor geslacht en ervaring. Met betrekking tot vooropleiding is er wel significant effect ($p = 0.02$) dat negatief blijkt te zijn: de evaluatiepraktijk van leerkrachten die in het bezit zijn van een masterdiploma zullen in vergelijking met leerkrachten bij wie het hoogste diploma van bachelorniveau is, in mindere mate gebruik maken van een semi-geïntegreerde evaluatiepraktijk.

In model 3b gaan we voor de secundaire scholen na of de onderwijsvorm waarin leerkrachten les geven een effect heeft op de mate waarin ze inzetten op een semi-geïntegreerde evaluatiepraktijk. We stellen geen significant effect vast voor leerkrachten die les geven in het bso ($p = 0.14$). Het effect van vooropleiding verdwijnt in model 3b na het controleren voor onderwijsvorm.

Het negatieve significante effect van leerlingencompositie blijft behouden in zowel model 3a als 3b.

Model 4: procesvariabelen (schoolniveau)

In model 4 wordt verder gewerkt op model 3a na het weglaten van de inputvariabelen op leerkrachtniveau (geslacht, ervaring en vooropleiding). Aan de hand van het vierde model gaan we het effect van de procesvariabelen op schoolniveau na: het beleidsvoerend vermogen van de school inzake leerlingenevaluatie, de wijze waarop leerlingen gegroepeerd worden, en de frequentie van klassenraden en leerlingencontacten. We stellen voor geen enkele van deze procesvariabelen op schoolniveau een significant effect vast.

Het negatieve, significante effect van leerlingencompositie blijft behouden ($p = 0.02$).

Model 5: procesvariabelen (leerkrachtniveau)

In model 5 voegen we de procesvariabelen op het niveau van de leerkracht toe. We laten de niet-significante procesvariabelen op schoolniveau weg en nemen de leerkrachtvariabelen geslacht, ervaring en vooropleiding terug op in model 5. Aangezien het om heel wat procesvariabelen gaat, doen we dit stapsgewijs: model 5a focust op de variabelen die verwijzen naar de praktijk van leerkrachten, de overige variabelen die verwijzen naar de opvattingen van leerkrachten worden extra toegevoegd in model 5b.

In model 5a worden de variabelen mate van professionalisering, de doelmatigheidsbeleving en het bekwaamheidsgevoel oog te hebben voor diversiteit onder de leerlingen toegevoegd. We merken een positief significant effect op voor de mate van professionalisering ($p < 0.001$) en het bekwaamheidsgevoel oog te hebben voor diversiteit ($p < 0.001$). Leerkrachten die inzetten op professionalisering inzake evaluatiepraktijk zullen meer geneigd zijn om in hun evaluatiepraktijk ook

aan te sluiten bij het semi-geïntegreerde model dan leerkrachten die minder inzetten op professionalisering. De score stijgt met een waarde van 0.65. Leerkrachten die zich bekwaamer voelen in het opmerken van diversiteit onder hun leerlingen zullen, in vergelijking met leerkrachten die zich hier minder bekwaam in voelen, eveneens meer inzetten op het semi-geïntegreerde model voor evaluatiepraktijk. De score neemt toe met een waarde van 0.23.

Het negatieve significante effect van leerlingencompositie verdwijnt in model 5a.

In model 5b worden de opvattingen van de leerkrachten toegevoegd. We stellen niet-significante effecten vast voor de onderwijsopvattingen van leerkrachten (leerlinggericht en leerstofgericht) en de opvattingen met betrekking tot diversiteit. De opvattingen met betrekking tot evaluatie zijn niet significant.

Beide positieve significante effecten voor mate van professionalisering en mate waarin men zich bekwaam voelt oog te hebben voor diversiteit onder leerlingen blijft behouden na het controleren voor de opvattingen van leerkrachten (telkens $p < 0.001$). We stellen vast dat het effect van leerlingencompositie terug significant is in model 5b ($p = 0.04$), na het meenemen van de opvattingen van leerkrachten.

Tabel 5.5

Resultaten van de multilevelanalyses met 'semi-geïntegreerde evaluatiepraktijk' als afhankelijke

	Model 0	Model 1	Model 2a	Model 2b	Model 3a
FIXED PART					
Intercept	3.73 (0.07)	3.72 (0.13)	3.83 (0.11)	3.38 (0.25)	3.83 (0.11)
SCHOOL					
Ligging		0.12 (0.14)			
Net:					
GO		0.28 (0.17)			
OGO		-0.09 (0.30)			
Niveau		-0.14 (0.14)	-0.16 (0.14)	0.00 (0.00)	-0.07 (0.15)
Compositie			-0.70 (0.25)**	-0.76 (0.41)	-0.64 (0.25)*
Aanbod bso				0.28 (0.27)	
Beleidsvoerend vermogen					
Leerlingencontact					
Klassenraad					
Groepering					
LEERKRACHT					
Geslacht					0.17 (0.16)
Ervaring					0.00 (0.01)
Opleiding					-0.36 (0.16)*
Onderwijsvorm					
Professionalisering					
Doelmatigheid					
Oog voor diversiteit					
Onderwijsopv. leerling					
Onderwijsopv. leerstof					
Opvatting diversiteit					
Opv. evaluatie informeren					
Opv. evaluatie kritisch					
Opv. evaluatie gn impact					
RANDOM PART					
School	0.00 (0.04)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Leerkracht	3.57 (0.18)	3.55 (0.18)	3.53 (0.18)	3.49 (0.25)	3.51 (0.18)
N (scholen)	56	56	56	22	56
N (leerkrachten)	816	816	816	398	816

*Noot: Per cel wordt weergegeven: regressiecoëfficiënt (standaardfout); * $p < ,05$, ** $p < ,01$, *** $p < ,001$*

Tabel 5.5 vervolg

Resultaten van de multilevelanalyses met 'semi-geïntegreerde evaluatiepraktijk' als afhankelijke

	Model 3b	Model 4	Model 5a	Model 5b
FIXED PART				
Intercept	3.88 (0.13)	3.97 (0.17)	3.82 (0.10)	3.80 (0.11)
SCHOOL				
Ligging				
Net:				
GO				
OGO				
Niveau	0.00 (0.00)	-0.36 (0.20)	-0.14 (0.15)	-0.10 (0.15)
Compositie	-0.71 (0.34)*	-0.067 (0.29)*	-0.43 (0.24)	-0.49 (0.25)*
Aanbod bso				
Beleidsvoerend vermogen		-0.12 (0.24)		
Leerlingencontact		-0.16 (0.15)		
Klassenraad		0.16 (0.17)		
Groepering		-0.05 (0.16)		
LEERKRACHT				
Geslacht	0.02 (0.19)		0.27 (0.16)	0.24 (0.16)
Ervaring	-0.00 (0.01)		0.00 (0.00)	0.00 (0.01)
Opleiding	-0.34 (0.18)		-0.24 (0.16)	-0.22 (0.16)
Onderwijsvorm	-0.28 (0.19)			
Professionalisering			0.65 (0.14)***	0.68 (0.14)***
Doelmatigheid			0.08 (0.10)	0.10 (0.11)
Oog voor diversiteit			0.23 (0.06)***	0.24 (0.05)***
Onderwijsopv. leerling				-0.13 (0.20)
Onderwijsopv. leerstof				-0.06 (0.16)
Opvatting diversiteit				-0.04 (0.12)
Opv. evaluatie informeren				-0.10 (0.14)
Opv. evaluatie kritisch				0.02 (0.11)
Opv. evaluatie gn impact				-0.06 (0.10)
RANDOM PART				
School	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Leerkracht	3.53 (0.23)	3.44 (0.18)	3.21 (0.16)-	3.12 (0.16)
N (scholen)	29	47	56	56
N (leerkrachten)	493	706	814	814

*Noot: Per cel wordt weergegeven: regressiecoëfficiënt (standaardfout); *p < ,05, **p < ,01, ***p < ,001*

3.3.3. Geïntegreerde evaluatiepraktijk als afhankelijke variabele

Model 0: het nulmodel

Het nulmodel omvat een model met twee niveaus en een random intercept met 'geïntegreerde evaluatiepraktijk' als uitkomstvariabele. Het intercept representeert de gemiddelde score voor 'geïntegreerde evaluatiepraktijk' voor alle leerkrachten die de vragenlijst invulden van de scholen die deelnamen aan het onderzoek. Uit dit model blijkt dat de gemiddelde score 2.84 (SD = 0.09) eerder laag is, wat erop wijst dat de evaluatiepraktijk van leerkrachten weinig gekenmerkt wordt door een sterke integratie van evaluatie en instructie.

De *variance partitioning coefficient* (VPC) heeft een waarde van 0.014. Dit toont aan dat slechts 1,4% van de totale resterende variantie zich op schoolniveau bevindt.

Model 1: contextvariabelen

De schoolvariabelen ligging, onderwijsnet en onderwijsniveau werden toegevoegd als bijkomende verklarende variabelen in model 1. Geen van deze extra toegevoegde variabelen hebben een significant effect op het hanteren van een geïntegreerde evaluatiepraktijk door leerkrachten.

Model 2: inputvariabelen (schoolniveau)

Omwille van de hoeveelheid aan variabelen worden de niet-significante contextvariabelen, onderwijsnet en ligging, niet opgenomen in model 2. De variabele onderwijsniveau behouden we en de inputvariabelen op schoolniveau, leerlingencompositie en studieaanbod, worden toegevoegd. We maken een onderscheid tussen model 2a en model 2b waarbij studieaanbod (referentiecategorie: zonder aanbod bso) toegevoegd wordt in een afzonderlijk model omdat deze variabele enkel van toepassing is op de scholen voor secundair onderwijs. Hiermee willen we rekening houden met het feit dat MlWin *listwise deletion* toepast, waardoor alle data die betrekking heeft op het basisonderwijs als *missing* behandeld wordt. Om een vertekend beeld te voorkomen voegen we deze variabele toe in een afzonderlijk model 2b.

In model 2a heeft de schoolvariabele leerlingencompositie een negatief significant effect ($p < 0.001$) wat erop wijst dat hoe groter het aandeel hoge SES-leerlingen op een school is, hoe minder de evaluatiepraktijk van leerkrachten op die school gekenmerkt zal worden door een geïntegreerde evaluatiepraktijk. De score neemt af met een waarde van -1.08. In model 2a merken we op dat het onderwijsniveau significant wordt na toevoeging van leerlingencompositie in het model ($p = 0.015$). Het effect is negatief: leerkrachten in het secundair onderwijs maken minder gebruik van een geïntegreerde evaluatiepraktijk dan leerkrachten in het basisonderwijs. De score voor een leerkracht in het secundair onderwijs is 0.37 punten minder.

Model 2b laat zien dat studieaanbod een positief significant effect heeft ($p = 0.016$). Leerkrachten in een school met aanbod bso zullen meer gebruik maken van een geïntegreerde evaluatiepraktijk dan leerkrachten van scholen die geen bso-aanbod hebben. Het verschil in uitkomstscore bedraagt 0.73.

We stellen vast dat het effect van leerlingcompositie verdwijnt in model 2b na het controleren voor studieaanbod.

Model 3: inputvariabelen (leerkrachtniveau)

Voor model 3a worden de inputvariabelen op leerkrachtniveau, geslacht, ervaring en vooropleiding toegevoegd aan model 2a waarin de variabelen onderwijsniveau en leerlingcompositie behouden blijven omwille van hun significant effect. Geen van deze nieuw toegevoegde variabelen heeft een significant effect op de evaluatiepraktijk van leerkrachten. Het significant effect voor onderwijsniveau en leerlingcompositie blijft behouden (respectievelijk $p = 0.046$ en $p < 0.001$).

In model 3b gaan we voor de scholen secundair onderwijs na of de onderwijsvorm waarin leerkrachten les geven een effect heeft op de evaluatiepraktijk. We stellen een positief significant effect vast voor het bso ($p < 0.001$): de evaluatiepraktijk van leerkrachten die les geven in het bso zal vaker aansluiten bij het geïntegreerde model dan bij leerkrachten die geen lesopdracht in het bso hebben. Concreet zal de score voor leerkrachten die in het bso les geven met een waarde van 0.71 toenemen.

Model 4: procesvariabelen (schoolniveau)

In model 4 wordt verder gewerkt op model 3a waarbij de niet-significante inputvariabelen op leerkrachtniveau (geslacht, ervaring en vooropleiding) weggelaten worden. Aan de hand van het vierde model gaan we het effect van de procesvariabelen op schoolniveau na: het beleidsvoerend vermogen van de school inzake leerlingenevaluatie, de wijze waarop leerlingen gegroepeerd worden, en de frequentie van klassenraden en leerlingcontacten. We merken hierbij één significant effect op, voor beleidsvoerend vermogen ($p < 0.001$). Er is een klein positief effect voor het beleidsvoerend vermogen: indien het beleidsvoerend vermogen van de school met één punt toeneemt dan neemt de mate waarin de evaluatiepraktijk van een leerkracht van die school aansluit bij het geïntegreerde model toe met een score van 0.12.

Het effect van onderwijsniveau en leerlingcompositie is in model 4 nog steeds significant (respectievelijk $p = 0.041$ en $p = 0.001$).

Model 5: procesvariabelen (leerkrachtniveau)

In model 5 voegen we de procesvariabelen op het niveau van de leerkracht toe. We laten de niet-significante procesvariabelen op schoolniveau weg en nemen de leerkrachtvariabelen geslacht, ervaring en vooropleiding terug op. Aangezien het om heel wat procesvariabelen gaat, voegen we ze stapsgewijs toe: in model 5a worden de variabelen toegevoegd die verwijzen naar de praktijk van leerkrachten, de overige variabelen die verwijzen naar de opvattingen van leerkrachten worden meegenomen in model 5b.

In model 5a stellen we significanties op voor de mate van professionalisering ($p = 0.001$) en het bekwaamheidsgevoel oog te hebben voor diversiteit onder de leerlingen ($p = 0.005$). Beide effecten

zijn positief: leerkrachten die werken aan hun professionalisering inzake leerlingenevaluatie en leerkrachten die zich bekwaamer voelen in het opmerken van diversiteit onder hun leerlingen zullen meer geneigd zijn in te zetten op een geïntegreerde evaluatiepraktijk. De uitkomstcores zullen respectievelijk toenemen met een waarde van 0.54 en 0.19. Voor de doelmatigheidsbeleving van leerkrachten stellen we geen significant effect vast voor de mate waarin de leerkracht gebruik maakt van het geïntegreerde model van evaluatiepraktijk. Naast het feit dat de significante effecten voor leerlingcompositie en beleidsvoerend vermogen behouden blijven, zien we een positief significant effect voor geslacht ($p = 0.04$) waarbij mannen meer inzetten op geïntegreerde evaluatiepraktijk in vergelijking met hun vrouwelijke collega-leerkrachten.

In model 5b worden de opvattingen van de leerkrachten toegevoegd. We stellen niet-significante effecten vast voor de onderwijsopvattingen van leerkrachten (leerlinggericht en leerstofgericht) en de opvattingen met betrekking tot diversiteit. De opvattingen met betrekking tot evaluatie zijn eveneens niet significant voor de mate waarin men ervan overtuigd is dat evaluatie een informerende waarde heeft of geen impact heeft op instructie. Er is een positief significant effect merkbaar voor de opvattingen van leerkrachten die verwijzen naar de mate waarin men er een kritische houding op na houdt met betrekking tot evaluatie ($p = 0.004$). Indien de kritische houding met één punt toeneemt, dan neemt de score voor het aansluiten bij een geïntegreerde evaluatiepraktijk toe met een waarde van 0.37.

Het significante effect van leerlingcompositie, geslacht, beleidsvoerend vermogen, het oog hebben voor diversiteit blijft behouden in model 5b. We zien ook dat na het toevoegen van de opvattingen van leerkrachten in het model, de doelmatigheidsbeleving van leerkrachten een positief significant krijgt ($p = 0.009$): leerkrachten die het gevoel hebben dat ze in staat zijn om doelmatig te handelen, zullen meer geneigd zijn om de geïntegreerde evaluatiepraktijk te hanteren.

In model 5b werd, ten slotte, ook de geaggregeerde variabele voor professionalisering op schoolniveau toegevoegd aangezien we eerder al vaststelden dat de mate van professionalisering bij leerkrachten een significant heeft. Dit blijkt geen significant effect te hebben.

Tabel 5.6

Resultaten van de multilevelanalyses met 'geïntegreerde evaluatiepraktijk' als afhankelijke

	Model 0	Model 1	Model 2a	Model 2b	Model 3a
FIXED PART					
Intercept	2.84 (0.09)	2.85 (0.15)	3.06 (0.12)	1.92 (0.28)	3.05 (0.12)
SCHOOL					
Ligging		0.17 (0.16)			
Net:					
GO		0.37 (0.19)			
OGO		0.50 (0.34)			
Niveau		-0.30 (0.16)	-0.37 (0.15)**	0.00 (0.00)	-0.34 (0.17)*
Compositie			-1.08 (0.28)***	-0.89 (0.47)	-1.01 (0.29)***
Aanbod bso				0.73 (0.31)*	
Beleidsvoerend vermogen					
Leerlingencontact					
Klassenraad					
Groepering					
LEERKRACHT					
Geslacht					0.21 (0.18)
Ervaring					-0.00 (0.00)
Opleiding					-0.24 (0.18)
Onderwijsvorm					
Professionalisering					
Doelmatigheid					
Oog voor diversiteit					
Onderwijsopv. leerling					
Onderwijsopv. leerstof					
Opvatting diversiteit					
Opv. evaluatie informeren					
Opv. evaluatie kritisch					
Opv. evaluatie gn impact					
RANDOM PART					
School	0.07 (0.07)	0.00 (0.05)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Leerkracht	4.59 (0.23)	4.59 (0.23)	4.55 (0.23)	4.52 (0.32)	4.53 (0.23)
N (scholen)	56	56	56	22	56
N (leerkrachten)	816	816	816	398	816

*Noot: Per cel wordt weergegeven: regressiecoëfficiënt (standaardfout); * $p < ,05$, ** $p < ,01$, *** $p < ,001$*

Tabel 5.6 vervolg

Resultaten van de multilevelanalyses met 'geïntegreerde evaluatiepraktijk' als afhankelijke

	Model 3b	Model 4	Model 5a	Model 5b
FIXED PART				
Intercept	2.37 (0.17)	2.92 (0.19)	2.97 (0.12)	2.99 (0.13)
SCHOOL				
Ligging				
Net:				
GO				
OGO				
Niveau	0.00 (0.00)	-0.42 (0.21)*	-0.30 (0.17)	-0.32 (0.19)
Compositie	-0.98 (0.38)**	-1.11 (0.32)**	-0.90 (0.28)**	-0.80 (0.29)**
Aanbod bso				
Beleidsvoerend vermogen		0.12 (0.11)***	0.24 (0.10)*	0.24 (0.11)*
Leerlingencontact		-0.01 (0.17)		
Klassenraad		0.16 (0.19)		
Groepering		0.18 (0.18)		
LEERKRACHT				
Geslacht	0.24 (0.21)		0.37 (0.18)*	0.40 (0.19)*
Ervaring	0.01 (0.01)		-0.01 (0.01)	-0.01 (0.01)
Opleiding	-0.14 (0.20)		-0.12 (0.18)	-0.14 (0.18)
Onderwijsvorm	0.71 (0.20)***			
Professionalisering			0.54 (0.16)**	0.43 (0.18)*
Doelmatigheid			0.23 (0.12)	0.33 (0.12)**
Oog voor diversiteit			0.19 (0.07)**	0.16 (0.07)*
Onderwijsopv. leerling				-0.37 (0.23)
Onderwijsopv. leerstof				0.11 (0.19)
Opvatting diversiteit				0.17 (0.14)
Opv. evaluatie informeren				0.01 (0.17)
Opv. evaluatie kritisch				0.37 (0.13)**
Opv. evaluatie gn impact				0.05 (0.12)
Professionaliseringsbeleid				0.66 (0.45)
RANDOM PART				
School	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Leerkracht	4.52 (0.29)	4.33 (0.24)	4.14 (0.21)-	4.07 (0.21)
N (scholen)	29	47	56	56
N (leerkrachten)	493	675	777	761

Noot: Per cel wordt weergegeven: regressiecoëfficiënt (standaardfout); * $p < ,05$, ** $p < ,01$, *** $p < ,001$

3.4. Conclusies bij multilevelanalyses voor de evaluatiepraktijk van leerkrachten.

Op basis van de resultaten die de nulmodellen opleveren, stellen we vast dat leerkrachten met betrekking tot hun evaluatiepraktijk gemiddeld genomen het vaakst gebruik maken van het gescheiden model (gemiddelde score van 5.34), dit gevolgd door het semi-geïntegreerd model (gemiddeld 3.73) en slechts af en toe gebruik maken van het geïntegreerde model.

Verschillen tussen scholen onderling doen zich voornamelijk voor bij het gescheiden model (VPC = 6.3 %) en in veel mindere mate bij het geïntegreerde model (VPC = 1.4 %). Voor het semi-geïntegreerde model stellen we nauwelijks verschillen vast die verklaard kunnen worden door factoren op schoolniveau (VPC = 0.1 %).

Voor de drie afhankelijke variabelen zien we dat de variatie op leerkrachtniveau telkens veel groter is dan de variatie op schoolniveau wat erop neerkomt dat de praktijk van evaluatie veeleer een zaak is van de individuele leerkracht dan van de school.

Met betrekking tot **het gescheiden model** zagen we dat de frequentie waarmee leerlingcontacten georganiseerd worden op school een voorspeller is voor het hanteren van een gescheiden evaluatiepraktijk. Leerkrachten op scholen die twee of meer leerlingencontacten per schooljaar organiseren, lijken vaker gebruik te maken van een gescheiden evaluatiepraktijk. Dit is een opmerkelijke vaststelling omdat we hier eerder een negatief verband zouden verwachten gezien de betrokkenheid van leerlingen bij het evaluatieproces negatief samenhangt met het gescheiden model (zie hoofdstuk 5.1). Tegelijkertijd zien we dat er bij het geïntegreerde model, waar een sterke betrokkenheid van leerlingen kenmerkend is (volgens bevindingen in hoofdstuk 5.1.), geen significant effect is voor de frequentie waarmee leerlingencontacten ingericht worden op school. Een mogelijke verklaring voor dit op het eerste gezicht atypische effect zouden we kunnen vinden in de wijze waarop gesprekken met leerlingen over hun leervorderingen of leerlingcontacten precies ingevuld worden. In hoofdstuk 5.1. stelden we al vast dat leerkrachten die in hun praktijk op het gescheiden model vertrouwen, veel aandacht besteden aan het verantwoordingsaspect. We zouden kunnen verwachten dat dit verantwoordingsaspect een sterke nadruk legt op de leerling zelf: het wel/niet behalen van vooropgestelde doelstellingen wordt dan veeleer toegeschreven aan de leerling zelf in plaats van aan het voorafgaand onderwijsproces. Deze positieve samenhang zou ook kunnen liggen in het feit dat leerkrachten opteren voor een gescheiden model precies omdat de school leerlingcontacten organiseert. Frequente leerlingcontacten zouden er dan toe leiden dat leerkrachten de evaluatie en instructie gaan loskoppelen waarbij tijdens een leerlingcontact teruggeblikt wordt op de uitkomsten van de evaluatie.

Het positieve effect van de frequentie waarmee leerlingencontacten georganiseerd worden, wordt ondersteund door een tweede significant effect dat we vonden bij het gescheiden model: namelijk het positieve verband met de overtuiging van leerkrachten dat evaluatie een informerende waarde heeft. Door beide effecten in verband te brengen zouden we kunnen afleiden dat de leerlingcontacten ingevuld worden als communicatie in één richting, waarbij de leerkracht zijn/haar bevindingen met betrekking tot het leerproces van de lerende meedeelt aan de leerling. Het lijkt erop dat dit éénrichtingsgesprek gekenmerkt wordt door het aanspreken van de verantwoordelijkheid van de leerling.

Bij het gescheiden model stelden we ook een negatief significant effect vast voor de onderwijsvorm bso. Leerkrachten in scholen met een aanbod bso zijn minder geneigd om hun evaluatie-activiteiten te scheiden van instructieactiviteiten. We zouden dit kunnen verklaren door de aard van de vakken in deze onderwijsvorm: binnen praktijkgerichte vakken staan vaardigheden centraal en vindt de evaluatie/het bijsturen van deze vaardigheden meer plaats terwijl de leerlingen deze vaardigheden inoefenen.

Een laatste significant effect ten slotte werd gevonden voor het onderwijsniveau. Leerkrachten in het secundair onderwijs zullen meer inzetten op het gescheiden model dan leerkrachten in het basisonderwijs. Als we kijken naar de kenmerken van een gescheiden model dan lijkt dit een logisch verschil: leerlingen in het secundair onderwijs zijn ouder waardoor leerkrachten meer geneigd zijn om het verantwoordingsaspect te beklemtonen. Het effect van onderwijsniveau lijkt echter te verdwijnen na toevoeging van de opvattingen van leerkrachten waardoor het verschil tussen beide onderwijsniveaus misschien eerder gezocht moet worden in hoe de leerkrachten van het basis- en secundair onderwijs verschillen in hun opvattingen over evaluatie: leerkrachten in het secundair onderwijs zijn meer overtuigd van de informerende waarde van evaluatie dan de leerkrachten in het basisonderwijs.

Met betrekking tot het **semi-geïntegreerde model** voor evaluatiepraktijk stelden we vast dat op het niveau van de school enkel de leerlingencompositie een effect heeft op de mate waarin leerkrachten inzetten op een semi-geïntegreerde evaluatiepraktijk. Hoe meer het leerlingenpubliek van een school gekenmerkt wordt door een hoge SES, hoe minder leerkrachten op deze school een semi-geïntegreerde evaluatiepraktijk zullen aanwenden. Dit betekent dat hoe meer lage SES-leerlingen het leerlingenpubliek vertegenwoordigen, hoe meer leerkrachten gevolg zullen geven aan wat men kan vaststellen uit de evaluatie. Dit kan ertoe leiden dat de instructie afgestemd wordt op waar de lerende staat in zijn/haar leerproces en/of dat er aan de hand van feedback concrete handvatten aangereikt worden waarmee de lerende verder kan. Dit is een positieve vaststelling, zeker ten aanzien van leerlingen met een lagere SES. Het sluit bovendien ook aan bij de eerdere bevinding dat leerkrachten die het semi-geïntegreerde model hanteren ook meer inzetten op differentiatie in het algemeen. We kunnen hier ook uit afleiden dat een leerlingenpubliek dat vooral gekenmerkt wordt door een hoge SES, veel minder kan rekenen op instructieactiviteiten die afgestemd worden op de uitkomsten van de leerlingenevaluaties. Deze vaststelling kan op twee manieren geïnterpreteerd worden. Enerzijds kunnen leerkrachten het niet nodig vinden om de instructie nog af te stemmen op basis van de evaluaties voor deze groep van leerlingen omdat men (mede door de hoge verwachtingen die men heeft ten aanzien van deze leerlingen) ervan uitgaat dat deze leerlingen zelf een extra inspanning kunnen leveren om de beoogde doelen te bereiken. Deze redenering sluit aan bij het verantwoordingsgerichte aspect, waarbij deze hoge SES-leerlingen aangesproken worden op hun verantwoordelijkheid. Anderzijds zou het achterliggende mechanisme kunnen zijn dat leerkrachten verwachten dat de beoogde doelen al behaald zijn of nog behaald kunnen worden door de leerling zelf. Dit laatste zou betekenen dat differentiatie en uitdaging die verder gaat dan de beoogde doelen, specifiek voor hoge SES-leerlingen, weinig aandacht krijgen van leerkrachten.

Tegelijkertijd stellen we vast dat de opvattingen van leerkrachten niet zozeer bepalend zijn voor de mate waarin ze inzetten op een semi-geïntegreerde evaluatiepraktijk. Wel blijkt hun bekwaamheid een rol te spelen want er is een positieve samenhang met de mate waarin leerkrachten inzetten op professionalisering en de mate waarin ze in staat zijn om diversiteit onder leerlingen op te merken. Het effect van professionalisering zou verklaard kunnen worden door bijvoorbeeld het feit dat leerkrachten die aan de slag gaan met semi-geïntegreerde evaluatiepraktijk misschien wel specifiek op zoek gaan naar professionalisering om te leren hoe ze dit het beste kunnen doen. Tevens is de mate waarin men zichzelf bekwaam acht om de diversiteit onder leerlingen op te merken, positief gerelateerd aan de mate waarin men rekening houdt met de uitkomsten van evaluaties bij het inrichten van een instructieactiviteit. Dit lijkt logisch: al naargelang wat de uitkomst is van evaluaties van leerlingen, zal de afstemming er anders uitzien. Dit vergt van leerkrachten het vermogen om te

kunnen vaststellen wie welke leernoden of -behoeften heeft. Het in staat zijn om verschillen tussen leerlingen op te merken vormt eigenlijk een voorwaarde om er ook op te kunnen inspelen tijdens momenten van instructie.

Het semi-geïntegreerde model laat ten slotte ook een significant effect zien voor de vooropleiding; een effect dat verdwijnt na het toevoegen van de procesvariabelen. Leerkrachten met een masterdiploma zouden minder inzetten op een semi-geïntegreerde evaluatiepraktijk dan leerkrachten bij wie het bachelordiploma het hoogst behaalde diploma is. Mogelijks is dit onderwerp meer verweven in het curriculum van een professionele bacheloropleiding dan in de specifieke lerarenopleiding waardoor leerkrachten met een bacheloropleiding vaardiger zijn in het afstemmen van de instructie op de resultaten uit de evaluatie.

De vaststellingen met betrekking tot het semi-geïntegreerde model duiken in nog sterkere mate op bij het **geïntegreerde model** voor evaluatiepraktijk. Dit geldt voor het effect van de leerlingensamenstelling, de mate van professionalisering en de bekwaamheid om diversiteit onder leerlingen op te merken. Bijkomend stellen we vast dat de doelmatigheidsbeleving van leerkrachten een positief effect heeft. Of anders gezegd: de mate waarin ze ervan overtuigd zijn dat hun handelen in de onderwijspraktijk een positief effect heeft, hangt positief samen met de mate waarin men gebruik maakt van het geïntegreerde evaluatiemodel. Leerkrachten die ervan overtuigd zijn dat hun handelen er toe doet, zullen dus meer geneigd zijn om evaluatieactiviteiten te integreren in hun lespraktijk.

Tegelijkertijd blijkt ook het geslacht van de leerkracht een rol te spelen: mannelijke leerkrachten zouden meer gebruik maken van een geïntegreerde evaluatiepraktijk dan vrouwelijke leerkrachten. Dit effect houdt ook stand na controle voor onderwijsvorm. Deze bevinding sluit aan bij onderzoek van Lietaert en Consuegra (2015) die vaststelden dat mannelijke leerkrachten verschillen van hun vrouwelijk collega's op vlak van leerkrachtstijl. Mannen zouden hoger scoren dan vrouwen op vlak van autonomie-ondersteuning en zouden meer inzetten op het aanreiken van keuzes aan de leerlingen. Het geïntegreerde model voor evaluatie gaat uit van evaluatiepraktijken die veel meer afgestemd zijn op het tempo van de leerling, wat bijgevolg ook beter aansluit bij een leerkrachtstijl die autonomie bij leerlingen bevordert.

In vergelijking met het semi-geïntegreerde model stellen we bij het geïntegreerde model een effect vast voor de opvattingen van leerkrachten: hoe kritischer leerkrachten zijn ten aanzien van evaluatie, hoe meer ze zullen inzetten op een geïntegreerde evaluatiepraktijk. Dit is een logisch effect: deze leerkrachten zijn kritisch ten aanzien van evaluatie omdat ze bedenkingen hebben ten aanzien van het validiteitsaspect: ze menen dat toetsresultaten onnauwkeurig kunnen zijn, meetfouten kunnen bevatten en denken dat dit oneerlijk kan zijn voor sommige leerlingen. Dit maakt dat deze leerkrachten in hun evaluatiepraktijk des te meer gaan inzetten op evaluatie die sterk ingebed is in de lespraktijk en aansluit op differentiatie, precies om tegemoet te komen aan de negatieve uitvloeisels die evaluatie met zich mee kan brengen. Tegelijk verwijzen de items voor een kritische houding ten aanzien van evaluatie naar een beperkt gebruik van toetsresultaten: *'leerkrachten doen aan evaluatie maar maken weinig gebruik van de resultaten ervan'* en *'toetsresultaten worden geklasseerd en genegeerd'*. Leerkrachten die aansluiten bij een geïntegreerde evaluatiepraktijk zullen de (uitkomsten van) evaluaties maximaal aanwenden voor het onderwijsleerproces.

Opmerkelijk is dat er naast de leerlingensamenstelling op school nog een tweede schoolvariabele als voorspeller blijkt te fungeren voor de mate waarin leerkrachten een geïntegreerde evaluatiepraktijk

aanwenden: het beleidsvoerend vermogen van de school. Dit zou mogelijk verklaard kunnen worden door het feit dat binnen dit geïntegreerde model evalueren en onderwijzen heel nauw met elkaar verweven zijn. De evaluatiepraktijk volgens dit model is ingebed in een ruimere visie op onderwijs en is bijgevolg ook sterk ingebed in een visie op leren. Dit sluit aan bij de bevinding van het eerdere onderzoeksluik, namelijk het casestudieonderzoek van Ysenbaert et al. (2018) waaruit bleek dat de scholen waarin een evaluatiebeleid afgestemd is op een ruimere visie op onderwijs en leren én waarbij de evaluatiepraktijken van leerkrachten onderling sterk op elkaar gelijken, gekenmerkt worden door een sterk beleidsvoerend vermogen.

Bij het geïntegreerd model stelden we ten slotte een positief significant effect vast voor de onderwijsvorm bso. De evaluatiepraktijk van leerkrachten die les geven in het bso sluit meer aan bij het geïntegreerde model dan leerkrachten die in andere onderwijsniveaus les geven. Dit kunnen we opnieuw verklaren door de aard van de vakken in het bso: bij praktijkgerichte vakken staan vaardigheden centraal en vindt de evaluatie/het bijsturen van deze vaardigheden meer plaats terwijl de leerlingen deze vaardigheden oefenen.

BIJLAGEN

1. Vragenlijst voor leerkrachten

2. Vragenlijst voor directies

3. Achtergrondinfo vragenlijsten

4. Correlatiematrix voor variabelen

Bibliografie

- Ainscow, M. (2007). *Teacher development in responding to student diversity: The way ahead*. Paper presented at the Responding to student diversity. Teacher education and classroom practice. Proceedings of the International Conference on Teacher Education for Responding to Student Diversity.
- Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2006). The Wheel of Competency Assessment: Presenting Quality Criteria for Competency Assessment Programs. *Studies in Educational Evaluation, 32*, 153-170.
- Banks, J., & al., e. (2007). Teaching Diverse Learners. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing Teachers for a Changing World*. San Fransisco.
- Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. *International handbook of research on teachers' beliefs*, 284-300.
- Birenbaum, M. (2014). Conceptualizing assessment culture in school. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *The enabling power of assessment*. (Vol. 1, pp. 285-302). Dordrecht: Springer.
- Birenbaum, M. (2016). Assessment Culture Versus Testing Culture: The Impact on Assessment for Learning. In D. Laveault & L. Allal (Eds.), *Assessment for Learning: Meeting the Challenge of Implementation*. Springer, Cham.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74.
- Black, P., & Wiliam, D. (2005). Lessons from around the world: how policies, politics and cultures constrain and afford assessment practices. *The Curriculum Journal, 16*(2), 249-261.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability(formerly: Journal of Personnel Evaluation in Education), 21*(1), 5. doi:10.1007/s11092-008-9068-5
- Bonner, S. M. (2016). Teachers' Perceptions About Assessment: Competing Narratives. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment*. London: Routledge.
- Brookhart, S. M. (2007). Expanding views about formative classroom assessment: A review of the literature. In J. H. Mc Millan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 43-62). New York: Teachers College Press.
- Brown, G. T., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and teacher education, 27*(1), 210-220.
- Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*(3), 301-318.
- Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review, 24*(2), 205-249.
- Clotfelter, C., Ladd, H., & Vigdor, J. (2006). Teacher-student matching and the assessment of teacher effectiveness. *The journal of Human Resources, 41*, 778-820.
- Coubergs, C., Struyven, K., Vanthournout, G., & Engels, N. (2017). Measuring teachers' perceptions about differentiated instruction: The DI-Quest instrument and model. *Studies in Educational Evaluation, 53*, 41-54.
- Dann, R. (2014). Assessment as learning: blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice, 21*(2), 149-166.

- Delandshere, G., & Jones, J. H. (1999). Elementary teachers' beliefs about assessment in mathematics: A case of assessment paralysis. *Journal of Curriculum and Supervision*, 14(3), 216.
- Denessen, E. (2000). *Opvattingen over onderwijs. Leerstof- en leerlinggerichtheid in Nederland*. Leuven-Apeldoorn: Garant.
- Dierick, S., & Dochy, F. (2001). New lines in edometrics: New forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation*, 27(4), 307-329.
- Duncan, C., Noonan, B., Simon, M., Tierney, R., Forgette-Giroux, R., & Hachey, K. (2009). Comparing grading practices between secondary mathematics teachers in Ontario and Saskatchewan. *Unpublished Manuscript, Canadian Educational Research Association, Canadian Society for the Study of Education, University of Ottawa, Ottawa*.
- Duncan, C. R., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research*, 53(1).
- Earl, L. M. (2003). *Assessment As Learning: Using Classroom Assessment to Maximize Student Learning*. London: Sage Publications.
- Ertmer, P. A., & Newby, T. J. (2013). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly*, 26(2), 43-71.
- Falchikov, N. (2013). *Improving assessment through student involvement: practical solutions for aiding learning in higher and further education*. London: Routledge.
- Fischman, W., DiBara, J. A., & Gardner, H. (2006). Creating good education against the odds. *Cambridge Journal of Education*, 36(3), 383-398.
- Gipps, C. (2011). *Beyond Testing (Classic Edition): Towards a theory of educational assessment*: Routledge.
- Griffiths, T., Gore, J., & Ladwig, J. (2006). *Teachers' fundamental beliefs, commitment to reform and the quality of pedagogy*. Paper presented at the Proceedings Australian Association for Research in Education Annual Conference.
- Harris, L. R., & Brown, G. T. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, 16(3), 365-381.
- Hattie, J. (2003). *Teachers make a difference: What is the research evidence?* Paper presented at the Australian Council for Educational Research Annual conference on Building teacher quality, Australia.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112.
- Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, 38(1), 21-27.
- James, M. (2006). Assessment, teaching and theories of learning. In J. Gardner (Ed.), *Assessment and learning* (pp. 47-60). London: Sage.
- Kozma, R. B., & Voogt, J. (2003). *Technology, innovation, and educational change: a global perspective: a report of the Second Information Technology in Education Study, Module 2*: ISTE (Interntl Soc Tech Educ).
- Lee, V. E., & Loeb, S. (2000). School size in Chicago elementary schools: Effects on teachers' attitudes and students' achievement. *American Educational Research Journal*, 37(1), 3-31.
- Liettaert, S., Consuegra, E. (2015). Verschilt de leerkrachtstijl naargelang het gender van de leerkracht? In: D. Van Maele, N. Michalek, N. Engels, F. Laevers, K. Lombaerts, M. Van Houtte (Eds.), *Gender op school. Meer dan een jongens-meisjeskwesitie*, Chapt. 25, (218-225). Leuven: Lannoo Campus.
- McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, 95(4), 203-213.
- Moon, T. R. (2005). The role of assessment in differentiation. *Theory Into Practice*, 44(3), 226-233.

- Moon, T. R. (2016). Differentiated Instruction and Assessment: An Approach to Classroom Assessment in Conditions of Student Diversity. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 284-301). New York: Routledge.
- Nisbet, S., & Warren, E. (2000). Primary school teachers' beliefs relating to mathematics, teaching and assessing mathematics and factors that influence these beliefs. *Mathematics Teacher Education and Development*, 2(34-47).
- Nye, B., Konstantopoulos, S., & Hedges, L. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26, 237-257.
- OECD. (2009). *Creating Effective Teaching and Learning Environments: First Results from TALIS*. Retrieved from Paris:
- OECD. (2013). *Synergies for Better learning: an international perspective on evaluation and assessment*. Retrieved from Paris:
- Onderwijsinspectie. (2006). *Onderwijspegiel: Verslag over de toestand van het onderwijs schooljaar 2005-2006*. Retrieved from Brussel: <http://www.onderwijsinspectie.be/sites/default/files/atoms/files/2005-2006.pdf>
- Opre, D. (2015). Teachers' conceptions of assessment. *Procedia-Social and Behavioral Sciences*, 209, 229-233.
- Penninckx, M., Vanhoof, J., & Van Petegem, P. (2011). *Evaluatie in het Vlaamse onderwijs. Beleid en praktijk van leerling tot overheid*. Antwerpen-Apeldoorn: Garant.
- Philips, I., Seghers, M., Versteden, P., & Ysenbaert, J. (2013). Toolkit Competenties Nederlands Breed Evalueren Lager Onderwijs. Gent: Steunpunt Diversiteit & Leren, Leuven: Centrum voor Taal en Onderwijs, Brussel: Vlaams Ministerie van Onderwijs en Vorming. Retrieved 10 November 2016 from <https://onderwijs.vlaanderen.be/sites/default/files/atoms/files/Toolkit%20breed%20evalueren%20competenties%20Nederlands%20in%20het%20lager%20onderwijs.pdf>
- Portael, L. (2013). *Behalen leerkrachten en scholen een goed rapport? Een onderzoek naar leerlingenevaluatie in het Vlaams secundair onderwijs.*, Antwerpen, Antwerpen.
- Remesal, A. (2007). Educational reform and primary and secondary teachers' conceptions of assessment: the Spanish instance, building upon Black and Wiliam. *The Curriculum Journal*, 18(1), 27-38.
- Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and teacher education*, 27(2), 472-482.
- Robinson, M., Lloyd, A., & Rowe, J. (2008). The impact of leadership on student outcomes: an analysis of the differential effects of leadership types. *Educational Administration*, 44(5), 635-674.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston.
- Sach, E. (2015). An exploration of teachers' narratives: what are the facilitators and constraints which promote or inhibit 'good' formative assessment practices in schools? *Education 3-13*, 43(3), 322-335.
- Scheerens, J. (1990). school effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement*, 1(1), 61-80.
- Schunk, D. H. (2012). *Learning theories an educational perspective sixth edition*: Pearson.
- Segers, M., & Tillema, H. (2011). How do Dutch secondary teachers and students conceive the purpose of assessment? *Studies in Educational Evaluation*, 37(1), 49-54.
- Shepard, L. A. (2000). The Role of Assessment in a Learning Culture. *Educational Researcher*, 29(7), 4-14.
- Sluijsmans, D. M. A., & Struyven, K. (2014). Quality assurance in assessment: An introduction to this special issue. *Studies in Educational Evaluation*, 43, 1-4. doi:10.1016/j.stueduc.2014.08.003
- Spolsky, B. (2004). *Language Policy*. Cambridge: Cambridge University Press.
- Stevens, P. A. (2007). Exploring the importance of teachers' institutional structure on the development of teachers' standards of assessment in Belgium. *Sociology of Education*, 80(4), 314-329.
- Struyf, E. (2000). *Evalueren: een leerkans voor leraren en leerlingen?* (Vol. 26). Leuven: Acco.

- Tomlinson, C. A. (2005). Grading and differentiation: paradox or good practice? *Theory Into Practice*, 44(3), 262-269.
- Tomlinson, C. A., & Moon, T. R. (2013). Differentiation and classroom assessment. In J. H. Mc Millan (Ed.), *Research on Classroom Assessment*. London: Sage
- Troudi, S., Coombe, C., & Al-Hamliy, M. (2009). EFL teachers' views of English language assessment in higher education in the United Arab Emirates and Kuwait. *TESOL Quarterly*, 43(3), 546-555.
- Tyler, R. W. (2013). *Basic principles of curriculum and instruction*: University of Chicago press.
- UNESCO. (1994). *The Salamanca Statement and Framework for Action on Special Needs Education*. Salamanca, Spain: United Nations Educational Scientific and Cultural Organization.
- Van Avermaet, P., Van Houtte, M., & Van den Branden, K. (2011). Promoting Equity and Excellence in Education. An overview in K. Van den Branden, P. Van Avermaet, & M. Van Houtte (Eds.), *Equity and Excellence in Education. Towards Maximal Learning Opportunities for All Students* (pp. 1-20). New York; London: Routledge.
- Van den Berg, M. (2001). *Onderwijsbeleid op de evenwichtsbalk. Studie naar effecten van deregulering en autonomievergroting in het onderwijs*. Retrieved from <https://www.onderwijsraad.nl/upload/documents/publicaties/volledig/deregbau.pdf>
- Van Petegem, P., & Vanhoof, J. (2002). *Evaluatie op de testbank. Een handboek voor het ontwikkelen van alternatieve evaluatievormen*. Mechelen: Wolters Plantyn.
- Vandeyar, S., & Killen, R. (2007). Educators' conceptions and practice of classroom assessments in post-apartheid South Africa. *South African Journal of Education*, 27(1), 101-115.
- Vanhoof, J., Deneire, A., & Van Petegem, P. (2011). *Waar zit beleidsvoerend vermogen van in (ver)scholen?* Mechelen: Plantyn.
- Verhoeven, C., Devos, G., Bruylant, W., & Warmoes, V. (2002). *Evalueren in Vlaamse secundaire scholen. Een onderzoek bij leraren en directeurs*. Antwerpen-Apeldoorn: Garant.
- Ysenbaert, J., Van Avermaet, P., & Van Houtte, M. (2017). Literatuurstudie evaluatie en diversiteit. Steunpunt Onderwijsonderzoek, Gent.
- Ysenbaert, J., Van Avermaet, P., & Van Houtte, M. (2018). Evaluatie & diversiteit in het basis- en secundair onderwijs. Deelrapport: casestudieonderzoek. Steunpunt Onderwijsonderzoek, Gent.