



Dit project werd gefinancierd met de steun van de Europese Commissie. De verantwoordelijkheid voor deze publicatieligt uitsluitend bij de auteur; de Commissie kan niet aansprakelijk worden gesteld voor het gebruik van de informatie die erin is vervat.



Een overzicht van nationaal beleid rond formatief evalueren

Introductie

Onderwijsbeleid wordt geacht een indirect effect te hebben op de prestaties van leerlingen door de onderwijs- en evaluatiepraktijk van scholen te verbeteren (Kyriakides, Creemers, Antoniou, Demetriou, & Charalambous, 2015). Een van de doelstellingen van het Erasmus + KA3-project, getiteld "Bevordering van formatieve evaluatie: van theorie tot beleid en praktijk (FORMAS)", is het genereren van beleidsrichtlijnen die de formatieve evaluatie bevorderen, aangezien onderzoek suggereert dat formatieve evaluatiepraktijken een positief effect hebben op de ontwikkeling van de student (Creemers & Kyriakides, 2015; Hattie & Temperley, 2007; Herman, Osmundson, Ayala, Schneider, & Timms, 2006; Wiliam, Lee, Harrison, & Black, 2004). Concreet probeert het project beleidsmakers aan te moedigen het evaluatiebeleid te hervormen en ondersteuningsmechanismen voor leraren op te zetten voor een effectieve implementatie van de formatieve evaluatie. Hiertoe werd een kritische analyse uitgevoerd van het huidige nationale beleid van de vier deelnemende landen (Cyprus, België, Griekenland en Nederland) met betrekking tot de formatieve en summatieve evaluatie van de wiskundekennis leerlingen. In dit verslag worden het kader en de methodologie voor het uitvoeren van de gegevensverzameling en-analyse gepresenteerd. Vervolgens worden de belangrijkste resultaten van de analyse gepresenteerd. Analyses zullen uitgevoerd worden om overeenkomsten en verschillen tussen de vier landen te identificeren. Het verslag eindigt met een kritische bespreking van de resultaten in verband met de belangrijkste bevindingen van onderzoek naar formatieve evaluatie om de sterke en de beperkingen van het huidige evaluatiebeleid vast te stellen.

1. Het raamwerk voor de beleidsanalyse

De term beleid verwijst naar een actieprincipe die door een organisatie of een individu wordt goedgekeurd of wordt voorgesteld (Cohen & Hill, 2001; Kyriakides et al., 2015; Kyriakides, Georgiou, Creemers, Panayiotou, & Reynolds, 2018). In het onderwijs betekent dit dat de beleidsmakers naar verwachting specifieke acties zullen introduceren en aanmoedigen die de belanghebbenden in de school (bijvoorbeeld management teams, docenten, leerlingen en ouders) moeten ondernemen om het leren van leerlingen te bevorderen. Deze acties worden voorgesteld in educatieve wetten en/of officiële documenten gepubliceerd door de nationale/lokale bestuursorganen en definiëren verordeningen, alsmede, de rol van de verschillende belanghebbenden in de richting van de effectieve werking van scholen. Tegelijkertijd worden relevante mechanismen ontwikkeld om de doeltreffende uitvoering van de voorgestelde beleidsrichtsnoeren en-verordeningen te ondersteunen.

In dit project wordt het onderwijsbeleid met betrekking tot evaluatie van leerlingen in het leergebied wiskunde van het secundair onderwijs in de vier deelnemende landen onderzocht. Naast het erkennen van de rol van lerarenopleiding en leraars professionele ontwikkeling (TPD) en hoe deze de praktijk van leraren kunnen beïnvloeden (DeSimone et al., 2002; Guskey & Sparks, 2004), worden ook beleidslijnen besproken die verband houden met initiële en in-service opleiding rondom formatieve en summatieve evaluatie in de klas. Bovendien wordt erkend dat evaluatiemechanismen die de uitvoering van het beleid onderzoeken, ook noodzakelijk zijn, aangezien zij besluiten nemen voor de voortzetting, beëindiging of aanpassing van uitgevoerde beleidsmaatregelen op basis van hun effectiviteit (Scheerens, 2016). Daarom wordt in dit project nationale onderwijsbeleid met betrekking tot de volgende drie aspecten in aanmerking genomen: a) leerlingenevaluatie, b) lerarenopleiding en TPD in de evaluatie van leerlingen, en c) evaluatiemechanismen met betrekking tot het beleid inzake evaluatie en TPD.

Het eerste aspect richt zich op beleid dat verwijst naar de evaluatie van leerlingen in de klas. Concreet onderzochten we beleid dat verwijst naar het doel van de evaluatie (d.w.z. formatieve en summatieve doelen), evaluatietechnieken, rapportageprocedures aan de betrokken belanghebbenden, de verstrekking van ondersteunende middelen, en het bestaan (indien aanwezig) van de nationale evaluatiemechanismen. Het tweede aspect richt zich op de opleiding van leraren en de professionele

ontwikkeling inzake evaluatie. Zo hebben we gekeken naar beleid dat verwijst naar de opleiding in de evaluatie als onderdeel van de programma's van de studie leidt tot een afgestudeerde van postdoctorale onderwijs, in-service professionele ontwikkeling en de ondersteuning aangeboden aan scholen door adviesorganen voor de uitvoering van de evaluatie van leerlingen. Daarnaast werd samenwerking tussen docenten binnen en tussen scholen in relatie tot evaluatie onderzocht. Ten slotte verwijst het derde aspect naar het beleid dat relevant is voor de mechanismen die zijn ontwikkeld om het nationale beleid (inclusief beleidshervormingen) op het gebied van evaluatie en TPD te evalueren. Concreet onderzochten we eerst of er een evaluatiemechanisme aanwezig is. Wanneer een dergelijk mechanisme bestond, hebben we gekeken naar hoe vaak en in welke fase van de uitvoering van het beleid de evaluatie wordt uitgevoerd, de bronnen van de gebruikte gegevens en processen, de focus van de evaluatie, alsmede, hoe de resultaten worden gebruikt en voor wie.

Bovendien wordt voor elk van de drie bovengenoemde aspecten rekening gehouden met zowel de kwantitatieve als de kwalitatieve kenmerken van het beleid. Dit is gedaan aan de hand van het dynamische model van de educatieve effectiviteit (Creemers & Kyriakides, 2008). Dit model erkent leerlingevaluatie als een factor die een impact op de leerresultaten heeft. Het model suggereert dat evaluatie kan worden gedefinieerd en gemeten in relatie tot vijf dimensies: frequentie, fase, focus, kwaliteit en differentiatie. Ten eerste werd het aantal relevante beleidsmaatregelen vastgesteld om de frequentie te onderzoeken die elk aspect in elk onderwijssysteem wordt aangepakt. Ten tweede, de focus van elk beleid werd onderzocht door te kijken naar de specificiteit ervan, alsmede, het doel of doeleinden die het beoogt te dienen. Ten derde werd de fase waarin het beleid plaatsvindt onderzocht. Op deze manier was het mogelijk om te bepalen wanneer een beleid wordt vastgesteld en of en wanneer op een gegeven moment is herzien, aangepast of beëindigd. Ten vierde werd onderzocht of het beleid in overeenstemming is met de literatuur. Op die manier konden conclusies worden getrokken over de kwaliteit van een beleidsmaatregel. Ten slotte werd rekening gehouden met de mate waarin een beleidsmaatregel differentieert ten aanzien van de groep van docenten (d.w.z. nieuw benoemde) of leerlingen (d.w.z. speciaal onderwijs leerlingen). Het gebruik van het genoemde meetkader stelde ons in staat een meer omvattend beeld te krijgen van het bestaande beleid. Tegelijk hielp het ons om meer specifieke suggesties en beleidsrichtlijnen te verstrekken die vormende evaluatie kunnen bevorderen.

2. Methodologie

Twee verschillende gegevensbronnen over het nationaal beleid inzake evaluatie in het kader van de wiskunde in het secundair onderwijs werden verzameld via: **(a) inhoudelijke analyse van beleidsdocumenten en (b) interviews met beleidsmakers die een sleutelrol spelen bij het bevorderen van evaluatiebeleid.** Op deze manier konden we niet alleen het officiële beleid meten zoals het is vermeld in documenten en richtsnoeren die naar scholen worden gestuurd, maar ook de acties van beleidsmakers in hun poging om de uitvoering van het nationale beleid op schoolniveau te bevorderen.

2.1. *Inhoudsanalyse van beleidsdocumenten*

Er werd een inhoudsanalyse uitgevoerd van alle officiële educatieve documenten van elk land (dat wil zeggen Cyprus, Griekenland, België en Nederland) die verordeningen en richtlijnen omvatten over aspecten die verband houden met: a) Leerlingenevaluatie, b) lerarenopleiding en TPD in evaluatie en c) evaluatiemechanismen met betrekking tot het beleid inzake evaluatie en TPD. De specifieke dimensies die voor elk aspect zijn onderzocht, zijn hierboven gepresenteerd en worden ook weergegeven in Tabel 1.

Voor elk aspect werd een profiel opgesteld door het bestuderen van de onderwijs wetten en alle voorschriften en richtlijnen die aan scholen worden meegedeeld (Zie het beleidsprofiel van het land op https://ucy.ac.cy/formas/documents/Deliverable_2.1_Analysis_of_Policies/Deliverable_2.1_Policy_Profile.pdf). De suggesties/voorschriften werden gedocumenteerd en geëvalueerd om scores per dimensie te genereren. Specifiek betekende dit dat elk onderzoeksteam elke beleidsmaatregel moest analyseren op de volgende aspecten: (a) de mate waarin scholen werden aangemoedigd om relevante acties te ondernemen, of moesten doen, of zelfs verantwoordelijk werden gehouden; b de belanghebbenden (d.w.z. secundair en/of primair onderwijs); c de aan de orde zijnde onderwerpen (d.w.z. alleen wiskunde en/of andere onderwerpen) (d) de mate waarin zij in overeenstemming was met de literatuur; e de mate waarin er enige vorm van differentiatie in de suggestie en de desbetreffende acties die daaruit zijn voortgekomen; en (f) de in eerste instantie vastgestelde termijn en de wijzigingen daarna. Hoe vaak een beleidsmaatregel per aspect voorkwam, werd ook in achtning genomen, omdat dit

een indicatie geeft van het gegeven belang aan een bepaalde beleidsmaatregel. Wat het evaluatiebeleid op deze beleidsmaatregelen in het onderwijs betreft, zijn de standpunten van elk land team over de nationale evaluatiemechanismen die in elk land bestaan, onderzocht. Alle partners van de universitaire instellingen en de ministeries werden geacht hun mening te geven en consensus te bereiken. Op basis van de gemeenschappelijke standpunten die zijn afgeleid, de zeven vragen opgenomen in het profiel werden beantwoord. De scores voor elke dimensie van elk aspect zijn ingevoerd in een SPSS-gegevensbestand. In het SPSS bestand werden zo veel kolommen als soorten acties gemaakt, om de frequentie dimensie te meten (Zie Excel-bestand gebruikt voor het invoeren van gegevens uit het beleid van het land profiel op <https://ucy.ac.cy/formas/en/resources> met 1-2 voorbeelden per land). De ingevulde beleidsdocumentenprofielen per land zijn beschikbaar in hun oorspronkelijke vorm op de website van ons project (zie <https://ucy.ac.cy/formas/en/resources>). Vervolgens wordt de methodologie voor het uitvoeren van de interviews met beleidsmakers gepresenteerd.

Tabel 1. De specifieke dimensies die onderzocht werden voor elk aspect van national onderwijsbeleid van de deelnemende landen

Aspecten van national onderwijsbeleid	Onderzochte dimensies
<i>Evaluatie</i>	
1. Doelen	
2. Technieken	✓ Frequentie
3. Rapportage van resultaten	
4. Aanwezigheid van middelen	✓ Focus 1 (Geen duidelijk beleid/ondersteuning/vereisten/verwachtingen)
5. Nationale evaluatiemechanismen (indien aanwezig)	✓ Focus 2 (Enkel voor secundair Onderwijs of voor basis- en secundair of anders)
<i>Training in evaluatie (opleiding en on-the-job) en samenwerking tussen leraren</i>	
6. Opleiding door hogescholen en universiteiten die tot grauaat of postgraduate diploma's leiden.	✓ Focus 3 (Enkel wiskunde of ook voor andere onderwerpen)
7. Professionele ontwikkeling: nascholingen en cursussen, inclusief mentorsystemen voor nieuw aangestelde leraren	✓ Fase (Periode van inzetten/evaluatie en aanpassingen van het beleid) ✓ Kwaliteit (cf. de literatuur)
8. Ondersteuning door adviserende organen	
9. Samenwerking rond evaluatie binnen scholen (inclusief team teaching, collegiale visitaties etc.)	✓ Differentiatie (met betrekking tot een groep leerlingen of leraren)
10. Stimuleren van netwerken tussen scholen	
11. Evaluatiemechanismen inzake het evaluatiebeleid en professionele ontwikkeling van leraren	

2.2. *Interviews met beleidsmakers*

Er werden semi-gestructureerde interviews gevoerd met beleidsmakers en andere stakeholders die belangrijke posities bekleden met betrekking tot de bevordering van het nationaal beleid rond leerlingenevaluatie op schoolniveau. In sommige landen (d.w.z. Nederland en België) zijn verschillende actoren uit verschillende niveaus van het onderwijssysteem verantwoordelijk voor de bevordering van specifieke aspecten van het evaluatiebeleid, terwijl in meer gecentraliseerde systemen, zoals Cyprus en Griekenland, een specifieke groep van ambtenaren (dat wil zeggen officieren van het ministerie van onderwijs) in staat waren om alle vragen te beantwoorden. Voor elk aspect van het beleid, verzamelden we interviewgegevens van ten minste drie beleidsmakers uit elk land, die werden geselecteerd met behulp van een “purposive sampling”.

In eerste instantie werden open vragen met betrekking tot de nationale beleidsmaatregelen en acties in verband met leerlingenevaluatie ontworpen samen met een checklist die ons in staat stelde om gegevens te genereren over elke dimensie en het aspect vermeld in tabel 1. De ontwikkelde interviewleidraad (zie interview gids op https://ucy.ac.cy/formas/documents/Deliverable_2.1_Analysis_of_Policies/Deliverable_2.1_Interview_Guide.pdf) omvatte 11 vragen, die elk werd gevolgd door een aantal follow-up en verdiepende vragen. Deze structuur zorgde ervoor dat alle genoemde aspecten van het onderzoek aan bod kwamen. Er werden ook richtlijnen ontwikkeld voor het analyseren van de interviewgegevens (Zie richtlijnen voor het analyseren van interview gegevens op https://ucy.ac.cy/formas/documents/Deliverable_2.1_Analysis_of_Policies/Deliverable_2.1_Guidelines_for_Analysing_Interview_Data.pdf). Deze richtlijnen werden voor elke geïnterviewde geanalyseerd, waaronder documentatie van de scores voor elk aspect van het nationale beleid. Om de anonimiteit van de geïnterviewden te garanderen, zijn alleen de documentatieformulieren in hun oorspronkelijke vorm en niet de transcripties van de interviews gerapporteerd. Een beschrijving van de te evalueren aspecten en dimensies kan worden gestrekt, inclusief voorbeelden. Daarnaast, de gedetailleerde instructies op de codering van de gegevens van het gesprek kunnen ook worden verstrekt indien nodig.

De volgende sectie presenteert de belangrijkste bevindingen van de data-analyse van beide beleidsdocumenten en de analyse van de interviews met de beleidsmakers in elk land met betrekking tot de aspecten en dimensies onderzocht.

3. Belangrijkste resultaten

De gegevens uit de analyse van beleidsdocumenten en de interviews van de beleidsmakers in elk land met elkaar vergeleken. De vier analyses toonden aan dat in elk land gegevens uit de analyse van de beleidsdocumenten analyse overeenkomen met de gegevens die uit de interview gegevens worden opgewekt. Zo werd steun voor de interne validiteit van deze studie verstrekt. Deze sectie bevat de resultaten van de analyse over de landen heen landen, welke gebaseerd zijn op zowel de inhoudelijke analyse van de beleidsdocumenten als de analyse de interviews met beleidsmakers.

3.1. Leerlingenevaluatie

Het FORMAS-project richt zich op verhogen van effectieve evaluatie van leerlingen in de wiskundeklas en is gericht op het ondersteunen van leraren uit het secundair onderwijs om te evalueren op manieren die het leren van leerlingen ondersteunen. Dus, het eerste onderzochte aspect verwijst naar het nationale beleid inzake leerlingenevaluatie. De relevante aspecten die hierbij zijn onderzocht zijn a) het doel van de evaluatie (dwz diagnostische, formatieve, summatieve evaluatie), b) evaluatietechnieken, c) rapportageprocedures aan de betrokken belanghebbenden, d) het verstrekken van ondersteunende middelen, en e) het bestaan (indien van toepassing) van nationale evaluatiemechanismen van beleidsmaatregelen in het onderwijs.

3.1.1. Het doel van evaluatie

Leerlingevaluatie kan een verscheidenheid van doeleinden dienen (Broadfoot, 1992; Brookhart, 2003; Gipps, 1994; Pellegrino et al., 2001; Torrance & Pryor, 1998). Op basis van de literatuur blijken summatieve en formatieve doeleinden de belangrijkste doeleinden. Summatieve evaluatie wordt gebruikt voor een eindevaluatie van een leerling op een systematische manier (DES/WO, 1988). Evalueren voor dit doeleinde is gericht op het beschrijven van wat de leerling heeft bereikt en wanneer

om op deze manier vergelijkingen te maken op basis van het niveau van de prestaties van de leerling. Aan de andere kant, wordt formatieve evaluatie gebruikt om de leerbehoeften van leerlingen te identificeren en gerichte ondersteuning te bieden om hen te helpen verbeteren. Zoals hierboven vermeld, blijken de formatieve evaluatiepraktijken een positief effect te hebben op de ontwikkeling van leerlingen (Hattie & Timperley, 2007; van de Herman, Osmundson, Ayala, Schneider, & Timms, 2006; Wiliam, Lee, Harrison, & zwart, 2004). Bovendien heeft onderzoek, tot nu toe, ook aangetoond dat het bereiken van beide doeleinden met één enkel mechanisme niet haalbaar is (Harlen & James, 1997; Zwart & Wiliam, 1998; Kyriakides & Campbell, 2003; Kyriakides, Demetriou, & Charalambous, 2006). Daarom wordt onderzocht of het onderwijsbeleid ondersteuning of bevordering van de summatieve en/of het formatieve doeleinden van evalueren in de wiskundeklas effectief is.

Om dit te doen, hebben we gekeken naar beleidsmaatregelen die ook de doeleinden van leerlingenevaluaties noemen, dit in alle vier deelnemende landen aan het FORMAS-project. Uit de analyse van beleidsdocumenten blijkt dat er in sommige documenten in de vier landen geen expliciete verwijzing is naar het doel van de evaluatie en dat sommige beleidsaanbevelingen, hoewel niet officieel vermeld, kunnen worden toegepast voor het bereiken van beide doeleinden. Bijna alle beleidslijnen zijn in overeenstemming met de literatuur, verwijzend naar het belang van een goede evaluatiepraktijk aangezien deze in wetenschappelijke literatuur wordt erkend. In Cyprus en Griekenland, is er een vereiste voor summatieve testen aan het eind van elk schooljaar die de voortgang van leerlingen evalueert. Dergelijke summatieve eindexamens vind je ook in de bovenbouw van het secundair onderwijs in Nederland. In dit land zorgen deze eindexamens wel voor verantwoordelijkheid, want als blijkt dat de resultaten niet voldoen aan de verwachtingen, worden scholen ondervraagd over de oorzaken daarvan en wordt gevraagd welke acties zij zullen ondernemen om de voortgang van de leerling te verbeteren. Aan de andere kant, in België, is de klassenraad het centrale evaluatie orgaan in het secundair onderwijs. Aan het eind van het schooljaar beslist de klassenraad of een leerling voldoende de doelstellingen van het onderwijsplan heeft bereikt om verder te gaan naar het volgende schooljaar op basis van alle evaluatiegegevens die gedurende het hele schooljaar zijn verzameld.

Bovendien, in alle vier landen is er beleid dat het gebruik van evaluatie van leerlingen aanmoedigt die leerlingen helpt hun vooruitgang te identificeren zodat zij inspanningen kunnen leveren

om te verbeteren. Echter, de meeste beleidslijnen die verwijzen naar formatieve doelen van evaluatie zijn niet verplicht te worden uitgevoerd door scholen, maar worden vermeld als suggesties die goed zijn om door te voeren. Specifieke acties die door de school, de docent of de leerling ondernomen moeten worden, komen in geen enkele beleidsmaatregel aan de orde. Dit wordt ook bevestigd door beleidsmakers in alle vier landen. Aan de andere kant, beleid dat verwijst naar summatieve evaluatie is wel aanwezig en houdt wel specifieke eisen en acties in. Geïnterviewden uit zowel Griekenland en Cyprus benadrukten dat, hoewel een aantal initiatieven ter bevordering van formatieve evaluatie kunnen worden geïdentificeerd, alleen summatieve evaluatiepraktijken officieel zijn vastgelegd. In Nederland en België, waar een autonoom systeem is opgezet, worden docenten aangemoedigd om hun leerlingen op een formatieve manier te evalueren, maar ze worden hier niet verantwoordelijk voor gehouden.

Andere opvallendheden zijn dat bijna alle beleidsmaatregelen betrekking hebben op zowel het primair als het secundair onderwijs. Bovendien, in geen van de beleidsmaatregelen wordt er specifiek gerefereerd aan het wiskundevak. Dit geldt voor alle aspecten die in de beleidsanalyse zijn onderzocht en deze bevinding zal verder worden besproken in het laatste deel van de opmerkingen.

Op basis van het bovenstaande kan de vraag worden gesteld of scholen evaluaties die het leren bevorderen wel kunnen uitvoeren of dat niet vermeden kan worden dat de school nadruk legt op de summatieve doeleinden van evalueren. Aangezien summatieve evaluaties meer lijken vast te liggen in beleidsrichtlijnen, is het niet verrassend dat evalueren voor formatieve doeleinden worden genegeerd. Zoals vermeld door een belanghebbende in Nederland, de eindexamens beïnvloeden andere evaluatiepraktijken van leraren, omdat scholen zich lijken aan te passen aan de vorm en de inhoud van deze examens. Soortgelijke opmerkingen werden ook gemaakt door beleidsmakers in alle andere landen. Uit de analyse van zowel beleidsdocumenten als interviews blijkt dat de verschillen tussen de twee doeleinden van de evaluatie niet naar behoren worden erkend. Uit onderzoek blijkt juist dat elk doel een heel andere benadering van evaluatie vereist. Bijvoorbeeld, de te gebruiken instrumenten voor evaluatie, de administratie van een evaluatie, en rapportage van evaluatiegegevens worden grotendeels bepaald door het doel van de evaluatie zelf. Daarnaast blijkt dat je niet beide doeleinden kunt bereiken met één evaluatiemethode. Het ontwikkelen van beleidsmaatregelen die specifiek het doel van de evaluatie bepalen, is dus noodzakelijk om de doeltreffende uitvoering ervan te waarborgen.

3.1.2. Evaluatietechnieken

Evaluatietechnieken hebben betrekking op de evaluatiemethoden die worden gehanteerd om het leren van leerlingen te beoordelen (bijv. schriftelijke evaluatie, mondelinge evaluatie, evaluatie door demonstratie). Evaluatietechnieken hebben een belangrijke rol bij het waarborgen van de kwaliteit en effectiviteit van de evaluatie, omdat ze meestal een invloed hebben op hoe en wat leerlingen leren. In de huidige literatuur wordt erkend dat er een verscheidenheid aan evaluatietechnieken moet worden gebruikt, aangezien leren multidimensioneel is en niet adequaat kan worden gemeten door een enkele techniek (Brookhart, 2003; Gipps, 1994).

Onderzoek naar beleidsmaatregelen die verwijzen naar specifieke evaluatietechnieken, wijst uit dat alle landen over het algemeen alleen beleidsmaatregelen hebben die niet specifiek verwijzen naar geschikte evaluatietechnieken. Bijna de helft van de genoemde evaluatietechnieken (d.w.z. 12 van de 23) worden vermeld als suggesties en zijn niet verplicht om te worden gebruikt. Bovendien wordt het doeleinde van de evaluatietechniek (formatief of summatief) niet vooraf bepaald. In Nederland werd bijvoorbeeld een Diagnostische Tussentijdse Toets (DTT) ontwikkeld om te worden afgenomen aan het einde van de onderbouw van het secundair onderwijs voor de vakken Nederlands, Engels en wiskunde. Deze toets was diagnostisch bedoeld en dit betekent dat deze toets diende om zo specifieke informatie te verstrekken over het leerproces aan de leerling en de docent. Echter, de beleidsmaatregel van de DTT is niet gecontinueerd (dat wil zeggen het werd slechts eenmaal toegediend in 2017). Op basis van de informatie die door de beleidsmakers tijdens de interviews wordt verstrekt, is het instrument moeilijk inzetbaar voor formatieve doeleinden, omdat het moment van de afname van de DIT samenvalt met een zogenaamde “high-stakes” beslissing (d.w.z. of de leerling kan doorstromen naar de hogere klassen van het secundair onderwijs). Tegelijkertijd is het wel noodzakelijk dat scholen een leerlingvolgsysteem hebben. Een voorbeeld hiervan is het Cito LVS, een leerlingvolgsysteem van Cito. Een dergelijk volgsysteem kan zowel ingezet worden voor formatieve en summatieve doeleinden, maar over het algemeen wordt deze vooral voor het laatstgenoemde gebruikt. Ook in Cyprus is geëxperimenteerd met het gebruik diagnostische testen gebruikt in het onderwijs, namelijk een diagnostische test in taal, wiskunde en wetenschap werd afgenomen in alle eerste klassen van het secundair onderwijs in september 2017. Echter, de test werd slechts eenmaal afgenomen en het gebruik ervan, namelijk het

identificeren van leerlingen met risico op uitval, was beperkt. Hoewel de gegevens beschikbaar waren, is het niet duidelijk of deze zijn gebruikt voor formatieve doeleinden. In België behoort het evaluatiebeleid tot de volledige autonomie van de school. Aan het eind van het technisch, kunst- en beroepsonderwijs zijn scholen echter verplicht een geïntegreerd examen te organiseren waarin specifieke competenties worden getoetst, maar het doel van deze evaluaties zijn niet duidelijk omschreven en lijkt in overeenstemming te zijn met het summatieve doel van de evaluatie. Hoewel authentieke evaluatietechnieken, zoals portfolios worden aangemoedigd als onderdeel van de formatieve evaluatie in Griekenland en Cyprus, verwijzen veel beleidsdocumenten (d.w.z. meer dan 60%) naar de eis om schriftelijke evaluaties met summatieve doeleinden te gebruiken.

Het bovenstaande laat zien dat docenten in alle vier de landen meestal vrij zijn om de evaluatietechnieken te kiezen die gebruikt moeten worden voor de evaluatie van het leren van leerlingen. Daarnaast blijkt dat evaluaties vaak gebaseerd zijn op het gebruik van één of een combinatie van de evaluatietechnieken. Zoals hierboven vermeld zijn niet alle technieken geschikt om alle leerdoelen te evalueren en suggereert de literatuur dat een verscheidenheid van evaluatietechnieken moeten worden gebruikt om de interne validiteit van de evaluatie te waarborgen. Daarom is het belangrijk dat beleidsmaatregelen meer specifieke aanwijzingen geven met betrekking tot het type en de verscheidenheid aan hulpmiddelen die moeten worden gebruikt voor de evaluatie van het leren van leerlingen. Ten slotte bevestigen zowel Cyprus als Nederland dat er pogingen worden gedaan om een evaluatie te gebruiken om de leerbehoeften van leerlingen te identificeren. Echter, deze pogingen hebben geen duidelijke formatieve oriëntatie en hebben geen continuïteit. Dit belemmert de positieve impact van een dergelijke evaluatie.

3.1.3. Rapporteren van resultaten aan stakeholders

De rapportage van resultaten overbruggt de kloof tussen het “afnemen” van een evaluatie, de analyse van de evaluatie en de interpretatie en het gebruik van evaluatieresultaten door de betrokkenen. De rapporteringsprocedures zorgen ervoor dat (a) de evaluatieresultaten tijdig en in begrijpelijke vorm beschikbaar zijn voor de verschillende gebruikers (Roeber, 2003) en (b) de continuïteit en de kwaliteit

van de leerervaring van leerlingen gewaarborgd wordt (BES, 2008). Rapportageprocedures bieden alle gebruikers kennis over de resultaten van de evaluatie, waardoor aanpassingen gemaakt kunnen worden ter ondersteuning van het onderwijs en leren.

Bij het analyseren van beleidslijnen die betrekking hebben op de rapportage van evaluatieresultaten, waren de verschillen tussen landen met een gecentraliseerd systeem (dwz Griekenland en Cyprus) en landen met een gedecentraliseerd systeem (dat wil zeggen België en Nederland) merkbaar. In Griekenland en Cyprus zijn wiskundedocenten verantwoordelijk om de resultaten te communiceren aan de beoogde gebruikers. Daarnaast zijn er specifieke beleidslijnen die bepalen wanneer en hoe de resultaten van de summatieve evaluatie aan de belanghebbenden worden meegedeeld. Bijvoorbeeld, in het beleid van Cyprus bepaalt de gebruikte schaal, het percentiel gegeven aan mondelinge en schriftelijke prestaties om het eindcijfer van de student te bepalen, alsmede, wanneer de resultaten worden meegedeeld aan leerlingen en ouders. Nogmaals, de nadruk ligt op summatieve evaluatie vermits geen beleid in beide landen verwijst naar rapportage voor formatieve doeleinden. In België en Nederland is de rapportage van de evaluatieresultaten een vereiste, maar is het doel van de rapportage niet verduidelijkt en scholen zijn vrij om te beslissen hoe dit zal gebeuren. Bijvoorbeeld, in Nederland is een mentor toegewezen voor elke klas en zal deze naar verwachting fungeren als de verbinding met de ouders. Hij/zij is in staat om informatie te rapporteren op basis van de resultaten van de student in verschillende vakken, hun vooruitgang en in relatie tot de specifieke klas in termen van prestaties en gedrag.

Op basis van het bovenstaande kan het ontbreken van een formatieve oriëntatie in de rapportage van evaluatieresultaten worden geïdentificeerd. Het feit dat het verstrekken van constructieve feedback aan alle belanghebbenden, en met name de student, een voorwaarde is voor de evaluatie van het leren van hulp, lijkt niet te worden erkend op beleidsniveau. Rapportage is de eerste stap om het gebruik van evaluatie-informatie mogelijk te maken, aangezien niemand kan handelen op gegevens die zij niet kennen. Bovendien moeten de methoden, de inhoud en de kwaliteit van deze mededeling worden aangepakt om ervoor te zorgen dat de resultaten zodanig worden gecommuniceerd dat aanpassingen mogelijk zijn om onderwijs en leren te ondersteunen.

3.1.4. Voorzien van middelen om leerlingevaluatie te ondersteunen

Vervolgens werden beleidsmaatregelen onderzocht die betrekking hebben op het verstrekken van (hulp)middelen die docenten kunnen ondersteunen bij de uitvoering van de evaluatie van leerlingen. De middelen in dit verslag hebben betrekking op elke vorm van ondersteuning aan leraren om hun praktijk in de evaluatie van leerlingen te ondersteunen. Deze middelen kunnen onder meer betrekking hebben op een opgavenbank of financiële steun. Een optimale verdeling in het gebruik van middelen is nodig om zowel doelmatige- als vermogensdoelstellingen te bereiken (Kyriakides, Creemers, & Charalambous, 2018).

De analyse toont aan dat naast Griekenland, de andere drie landen weinig of geen beleid hebben dat naar ondersteunende middelen verwijst. In het geval van Griekenland, kunnen een aantal centraal ontwikkelde middelen worden geïdentificeerd, maar deze worden alleen vermeld als suggesties. Hun doel is niet duidelijk omschreven en scholen zijn niet verplicht om ze te gebruiken. In Cyprus zijn prestatie-indicatoren per onderwerp beschikbaar en worden docenten geacht hun onderwijs op basis van hen te plannen, uit te voeren en te evalueren. Echter, alle beleidsmakers erkenden in de interviews dat dit beleid verdere verbetering nodig heeft, omdat docenten niet zeker zijn over hoe deze prestatie-indicatoren te gebruiken om hun onderwijs te helpen. Bovendien maakt het grote aantal geïdentificeerde indicatoren het voor leraren moeilijker om ze te gebruiken om hun onderwijs en evaluatie te plannen. Ook België heeft een beleid ontwikkeld dat verwijst naar evaluatie-indicatoren die gericht zijn op het helpen ontwikkelen van een duidelijkere evaluatie. In Nederland worden de DIT, evenals, de Cito LVS tests vermeld als bronnen die kunnen worden gebruikt om leraren te ondersteunen, maar opnieuw is het aan de scholen om te beslissen of ze te gebruiken en voor welk doel.

De bovenstaande resultaten laten zien dat docenten niet erg worden ondersteund om de evaluatie te gebruiken voor formatieve doeleinden. Gezien het feit dat de uitvoering van de evaluatie voor formatieve doeleinden moeilijker is, maakt de beperkte ondersteuning het nog moeilijker voor docenten om de evaluatie voor formatieve in plaats van summatieve doelen te gebruiken. Hoewel sommige ondersteunende mechanismen in alle vier landen kunnen worden geïdentificeerd, worden deze niet als eis gesteld en daarom is het aan de docenten zelf om te beslissen of zij van hen zullen gebruik maken

en hoe. Leerlingenevaluatie is een proces dat meerdere stappen en procedures impliceert en de verwachting is dat docenten in staat zijn en/of bereid zijn om ze correct uit te voeren. Op eigen initiatief is dat niet zo vanzelfsprekend. Het is duidelijk dat beleidsmaatregelen die voldoende ondersteuning bieden in de vorm van verschillende middelen (bijvoorbeeld financiële steun en evaluatie-instrumenten) noodzakelijk zijn om de docenten te ondersteunen bij het effectief formatief evalueren van de (wiskunde)kennis en -vaardigheden van hun leerlingen.

3.1.5. Nationale evaluatietechnieken

Ten slotte werd het beleid met betrekking tot nationale evaluatiemechanismen onderzocht. In Cyprus en Griekenland is er slechts één nationaal evaluatiemechanisme en deze heeft betrekking op de centrale middelbare schoolexamens. Dit zijn toelatingsexamens voor de universiteit en zijn een vereiste voor alle leerlingen van de zesde klas. Beleidsmakers in beide landen bevestigden dat de hoge inzet van deze examens ertoe leidt dat de docenten hun inspanningen gedurende het hele schooljaar concentreren op de vraag hoe zij hun leerlingen kunnen helpen de hoogst mogelijke prestaties te bereiken.

In België is het nationale evaluatieprogramma ('Peilingen') sinds 2002 het enige nationale evaluatieprogramma dat tot doel heeft betrouwbare informatie te verzamelen en te verstrekken over de validiteit en haalbaarheid van de eindtermen, over de resultaten van het schoolsysteem met betrekking tot deze eindtermen, en over de doeltreffendheid van de processen die in het huidige schoolsysteem worden gebruikt om deze resultaten te bereiken. Hoewel het een beleid is dat voor formatieve doeleinden kan worden gebruikt, is het geen vereiste. Anderzijds zijn er in Nederland een aantal landelijke evaluatiemechanismen in het leven geroepen. Ten eerste is sinds 2017 een beleid dat vereist dat de school informatie verzamelt over de kennis en vaardigheden van haar leerlingen vanaf het moment dat zij de school binnenkomen. De school moet deze informatie vergelijken met de verwachte ontwikkelingen. Deze informatie en vergelijking maken het mogelijk om het onderwijs aan te passen aan de onderwijsbehoeften van individuele leerlingen en/of groepen leerlingen op basis van hun capaciteiten. Wanneer leerlingen geen baat lijken te hebben bij hun opleiding, wordt van de school verwacht dat ze de reden hiervoor identificeren en een mogelijke verklaring geven. Zij beslissen wat er moet gebeuren om bepaalde achterstanden weg te werken en de leerlingen krijgen de begeleiding die

ze nodig hebben om het onderwijsprogramma adequaat te doorlopen. Voor leerlingen met een leerachterstand is het onderwijs zo gestructureerd dat er merkbare aandacht kan worden besteed aan het wegwerken van deze leerachterstand. Dit beleid dient duidelijk de formatieve doeleinden van evaluaties, aangezien het gebruik maakt van de resultaten van evaluaties om het onderwijs aan te passen en het leren van leerlingen te ondersteunen. Daarnaast is sinds 2017 elke school verplicht om een Programma van Toetsen en Afsluiting (PTA) en een examenreglement te ontwikkelen dat de relevante wetgeving van het land volgt. In dit document worden de evaluatieprocessen van de leerlingen in per school beschreven.

Op basis van de analyse van het beleid met betrekking tot het eerste onderzochte aspect, de evaluatie van leerlingen, komen we tot de conclusie dat de evaluatie van leerlingen in alle deelnemende landen als belangrijk wordt erkend. Van docenten wordt verwacht dat zij hun leerlingen evalueren voor zowel summatieve als formatieve doeleinden. Hoewel formatieve evaluatiepraktijken een positieve impact lijken te hebben op het leren van leerlingen (Kyriakides, 2005; Kyriakides & Creemers, 2008), worden formatieve evaluatiespraktijken echter niet als vereisten gesteld. Beleid dat verwijst naar het gebruik van evaluatietechnieken en het rapporteren van evaluatiesinformatie aan de beoogde gebruikers, houdt docenten alleen verantwoordelijk als het gaat om summatieve evaluatie. Rekening houdend met het gebrek aan ondersteunende middelen voor de evaluatie van leerlingen, worden vragen gesteld over de vraag of en hoe de formatieve evaluatiepraktijk wordt geïmplementeerd.

3.2. Lerarenopleiding en TPD in leerlingevaluatie

Het tweede aspect dat onderzocht is, is het beleid rondom de integratie van leerlingevaluatie in lerarenopleidingen en TPD. Lerarenopleidingen en professionele ontwikkeling worden beschouwd als belangrijke methoden om de effectiviteit van een school te verhogen (Darling-Hammond, 2000; Guskey, 2003; Smith & O'Day, 1991). Zo hebben we onderzoek gedaan naar beleid dat specifiek verwijst naar training in evalueren als onderdeel van de studieprogramma's die leiden tot een gediplomeerde of postdoctorale graad in het onderwijs, bijscholing en de ondersteuning die scholen krijgen aangeboden door adviesorganen voor de implementatie van leerlingevaluatie. Ook de

samenwerking van docenten binnen en tussen scholen met betrekking tot leerlingevaluaties, aangezien onderzoek suggereert dat professioneel leren van docenten niet alleen plaatsvindt in georganiseerde trainingen, maar ook door middel van een verscheidenheid aan minder formele interacties (d.w.z., professionele leergemeentes) (Richter et al., 2014).

3.2.1. Initiële opleiding van docenten (hoger onderwijs)

De initiële opleiding van docenten is de opleiding die wordt aangeboden aan toekomstige docenten als onderdeel van de studieprogramma's die leiden tot een gediplomeerde of postdoctorale graad in het onderwijs. Van de initiële opleiding wordt verwacht dat zij de docenten voldoende voorbereidt om hun complexe rol in een echte schoolomgeving effectief te vervullen (Darling-Hammond, 2006). Gezien de grote hoeveelheid tijd die docenten besteden aan evaluatie-gerelateerde activiteiten, wordt verwacht dat de opleiding van docenten vóór de indiensttreding onder meer de nodige opleiding in het evalueren van leerlingen zal bieden. Om deze reden is onderzocht op welke manier het officiële beleid eisen stelt aan of ondersteuning geeft aan de ontwikkeling van docenten in leerlingevaluatie voordat zij aan het werk gaan als docent.

Ten eerste was het opvallend dat geen van de vier landen een formeel beleid heeft dat de training in evalueren als voorwaarde stelt voor de aanstelling van docenten. Uit interviews met beleidsmakers in alle landen is gebleken dat onderwijsautoriteiten banden en gemeenschappelijke projecten hebben met instellingen voor hoger onderwijs, maar dat zij geen zeggenschap hebben over het curriculum of de inhoudelijke ontwikkeling van bachelor- of postdoctorale opleidingen die leiden tot een doctoraal of postdoctoraal diploma van een leraar. Dit roept vragen op over tot in hoeverre een docent het evalueren van de leerlingen als onderdeel op waarde schat, vooral wanneer zij net zijn aangesteld. Terwijl het evalueren van leerlingen in het takenpakket van de docent is opgenomen, is er geen beleid aanwezig dat vereist dat zij goed geïnformeerd en bekwaam zijn in het evalueren. Het zou daarom belangrijk kunnen worden geacht om van docenten te eisen dat zij een minimum aan evaluatievaardigheden hebben ontwikkeld voordat zij het beroep kunnen gaan uitoefenen. Het feit dat er voor de aanstelling

van docenten geen initiële opleiding in evalueren nodig is, benadrukt het belang van bijscholing in het goed kunnen evalueren van leerlingen.

3.2.2. Bijscholing en professionele ontwikkeling van de docent

TPD wordt beschouwd als een belangrijke manier waarmee de inhoudelijke kennis van docenten en de ontwikkeling van hun onderwijspraktijk naar een hoger niveau kunnen worden gebracht (Borko, 2004). Onderzoek suggereert dat docenten de kwaliteit van hun praktijk kunnen verbeteren door deel te nemen aan passende interventies en aan effectieve programma's voor professionele ontwikkeling (Creemers, Kyriakides, & Antoniou, 2013).

Als we kijken naar het beleid met betrekking tot TPD in leerlingevaluatie (d.w.z. bijscholingscursussen, mentorsystemen voor nieuw aangestelde docenten, enz.) zien we dat deze in alle vier de landen aangeboden worden, maar meestal alleen op basis van vrijwilligheid. Daarnaast is de deelname aan zulke TPD-mogelijkheden vaak niet gekoppeld aan de ontwikkeling van de loopbaan van de docent.

In Griekenland moesten in de jaren 2010 tot 2011 nieuw aangestelde docenten een introductiecursus volgen die onder meer een opleiding in leerlingevaluatie omvatte. Dit beleid werd echter stopgezet. In Cyprus moeten alle nieuw aangestelde docenten in het secundair onderwijs een opleidingsprogramma volgen om hun opleidingscapaciteit te ontwikkelen vóór hun aanstelling. Onder de aangeboden cursussen is ook een cursus gericht op het evalueren van leerlingen, maar deze is facultatief en niet alle docenten hoeven deze af te ronden. In 2018 legde het Cypriotische ministerie van Onderwijs en Cultuur meer nadruk op het goed evalueren van leerlingen en alle docenten van het voortgezet onderwijs werden verplicht een seminarie bij te wonen over authentieke evaluatiepraktijken georganiseerd door de coördinatoren van de vakken. De beleidsmakers benadrukten in hun interviews echter dat, hoewel dit een stap in de goede richting was, de duur en focus van deze training niet voldoende was. In België zijn alle scholen sinds 2009 verplicht om een jaarplan voor in-service training op te stellen. Dit plan houdt rekening met alle opleidingsinspanningen die gericht zijn op de ontwikkeling van de kennis, competenties en attitudes van de docent. Het evalueren van de leerlingen kan deel uitmaken van dit plan, maar dit is geen vereiste. In Nederland is in 2018 een beleid ten aanzien

van TPD als vereiste ingevoerd. Van docenten werd verwacht dat zij verantwoording afleggen over hun ontwikkeling in het onderwijs door relevante informatie in het Lerarenregister in te voeren. De invoering van het Lerarenregister is in datzelfde jaar echter stopgezet (zie <https://www.vo-raad.nl/themas/36/onderwerpen/330> voor meer informatie).

Zoals gezegd, is TPD een krachtig instrument om verbeteringen in het onderwijs te bevorderen (Kennedy, 2016). Uit de analyse van zowel beleidsdocumenten als interviews in alle vier de landen blijkt echter dat de meeste opleidingsmogelijkheden op vrijwillige basis worden aangeboden. Dit doet de vraag rijzen of alle docenten, en met name degenen die niet over de basisvaardigheden beschikken voor het evalueren, de nodige opleiding krijgen. Daarom wordt de noodzaak van een meer gestructureerd en gericht beleid op het gebied van de professionele ontwikkeling van docenten als noodzakelijk gezien. Bovendien is het zo dat zelfs wanneer er mogelijkheden voor professionele ontwikkeling worden geboden, deze meestal van korte duur zijn en geen continuïteit hebben, zoals in de meeste Europese landen het geval is. Uit onderzoek is gebleken dat de workshops voor professionele ontwikkeling op de werkvloer meestal niet aansluiten bij de bestaande praktijken, behoeften en prioriteiten van de deelnemers en niet op betrouwbare wijze leiden tot veranderingen in de docentpraktijk (Antoniou & Kyriakides, 2013; Loucks-Horsley et al., 1998). Daarom moet de duur van de programma's voor professionele ontwikkeling en de tijdsperiode waarover deze programma's verspreid zijn (Desimone, 2009). Daarbij moet opgemerkt dat beleidsmaatregelen ontwikkeld moeten worden die erop gericht zijn de evaluatiepraktijk te ondersteunen door middel van interventies voor professionele ontwikkeling. Ten slotte wordt betoogd dat om professionele ontwikkeling als effectief te beschouwen, het gestructureerd professioneel leren moet voorzien in gestructureerd professioneel leren dat leidt tot veranderingen in de praktijk van docenten en verbeteringen in de leerresultaten van leerlingen (Darling-Hammond, Hylar, & Gardner, 2017). Dit onderstreept het belang van het onderzoeken van de impact van professionele leerinitiatieven op zowel de lerarenpraktijk als het leren van leerlingen.

3.2.3. *Ondersteuning door schooladviesorganen aan docenten bij het evalueren van leerlingen*

Het volgende aspect dat werd onderzocht, was de ondersteuning die docenten krijgen bij het evalueren van leerlingen door schooladviesorganen. Op basis van de analyse van het nationale beleid was dit aspect niet van toepassing op Griekenland. In Cyprus wordt de officiële bijscholing van docenten voornamelijk aangeboden door het Pedagogisch Instituut van Cyprus. Het Pedagogisch Instituut van Cyprus biedt beroepsopleidingen aan door middel van een reeks facultatieve seminars. Sinds 2015 heeft het Cypriotische ministerie van Onderwijs en Cultuur een beleidsmaatregel over professioneel leren ingevoerd. Deze maatregel vereist dat alle scholen een plan voor professionele ontwikkeling op school ontwikkelen dat gebaseerd is op de specifieke behoeften van de docenten en rekening houdt met de specifieke kenmerken van de school. Elke school kiest haar eigen prioriteitsgebied en bedenkt acties voor de ontwikkeling van docenten in relatie tot de vastgestelde behoeften. Het programma Professioneel Leren wordt uitgevoerd met de steun van het Pedagogisch Instituut. In Nederland heeft het College voor Toetsen en Examens (CvTE) in 2015 een aantal scholen gefaciliteerd in het onderzoek naar de toegevoegde waarde van de Diagnostische Tussentijdse Toets (DTT). Later werd echter gesuggereerd dat het instrument ook voor formatieve doeleinden kan worden gebruikt. Het CvTE heeft leerlingen en docenten ondersteund door hen te helpen bij het in kaart brengen hoe de tool past in de cyclus van formatieve evaluatie en op welke manier de scholen de diagnose kunnen gebruiken om hun praktijk te verbeteren. Deze ondersteuning werd echter alleen aangeboden aan de 24 scholen die deelnamen aan de pilotstudie en werd vervolgens stopgezet. Zoals een beleidsmaker in Nederland verder uitlegde, bemoeilijkte het feit dat het instrument niet in de eerste plaats was ontworpen voor formatieve doeleinden, het formatieve gebruik ervan. Daarnaast hebben docenten recht op financiële steun (600 per fulltime docent) en 83 uur TPD en 83 uur per schooljaar. De uren en het budget worden alleen besteed aan activiteiten die door de begeleider worden toegewezen, maar alleen als de docent het daarmee eens is. Hierbij wordt niet specifiek genoemd dat de training gericht moet zijn op leerlingevaluatie. In België heeft elke onderwijskoepelorganisatie haar eigen Pedagogische begeleidingsdienst, die de scholen op verzoek professionele interne ondersteuning biedt. Scholen kunnen een beroep op hen doen voor pedagogisch en methodologisch advies (bv. innovatieprojecten, zelfevaluatieprojecten, ondersteunende initiatieven). Schooladviseurs werken in alle scholen voor de

in-diensttraining en ondersteuning van schooldirecteuren en schoolteams. Zij ondersteunen en stimuleren ook mechanismen voor kwaliteitsborging in de scholen. Alle drie beleidsmakers benadrukken in hun interviews de verschuiving van ondersteuning van individuele docenten naar ondersteuning op systeemniveau. Ze vermelden ook dat door de schooladviesdiensten verschillende in-service trainingen worden georganiseerd om het evaluatie- en rapportagebeleid voor docenten te ontwikkelen, gebaseerd op schooladviesinitiatieven en op de vraag van schoolleiders. Er is echter geen duidelijk beleid op dit gebied.

Op basis van het bovenstaande kan betoogd worden dat wanneer de beslissing over professionele ontwikkeling op de eerste school waar de docent werkt gebaseerd is en zonder hierbij de opleidingsbehoeften van de docenten te evalueren het meer mogelijk is dat het programma wordt aangepast aan wat de school belangrijk vindt en met de behoeften van de deelnemende docenten. In beleidsmaatregelen die het gebruik van schoolgebaseerde benaderingen van TPD ondersteunen, moet dus rekening gehouden worden met zowel (a) belangrijke contextuele kenmerken, en (b) de noodzakelijke kennisbasis die nodig is om ervoor te zorgen dat initiatieven gebaseerd zijn op onderzoek en dus waarschijnlijk een positieve bijdrage leveren aan de praktijk van de docenten.

3.2.4. Samenwerking tussen docenten binnen scholen en tussen scholen

"Samenwerking kan worden gedefinieerd als gezamenlijke interactie in de groep bij alle activiteiten die nodig zijn om een gezamenlijke taak uit te voeren" (Vangrieken, Dochy, Raes, & Kyndt, 2015, p. 23). Hoogwaardige samenwerking tussen docenten kan positieve effecten hebben op de kwaliteit van lesgeven van de docent en hierdoor op de prestatie van hun leerlingen. Hierbij wordt juist evaluatiegerichte samenwerking wordt door leerkrachten erkend als een zeer nuttig instrument om hun praktijk te verbeteren (Ronfeldt, Farmer, McQueen, & Grissom, 2015). Het onderzoeken van het aspect van de samenwerking tussen docenten binnen en tussen scholen op het gebied van leerlingevaluaties toont echter aan dat geen van de vier landen relevante vereisten heeft. In België en Nederland worden docenten en scholen aangemoedigd om samen te werken om hun onderwijspraktijk te verbeteren. Toch wordt dit niet als eis gesteld, maar vooral als een suggestie voor goede praktijken. Een voorbeeld van

het perspectief op wat een goede praktijk is in Nederland, werd in Nederland geanalyseerd in een case studie in 2005. Daarnaast is er sinds 2016 een beleid ten aanzien van het netwerken van scholen gericht op het faciliteren van het gebruik van formatieve evaluatie in de klas. In vijf netwerkbijeenkomsten wordt van scholen verwacht dat zij samenwerken aan zaken die te maken hebben met formatief evalueren. Het belangrijkste doel van deze bijeenkomsten is dat scholen van elkaar leren en in de periode tussen de bijeenkomsten door wordt er gedifferentieerde begeleiding en advies gegeven door Stichting Leerplan Ontwikkeling (SLO). In Griekenland worden docenten die lesgeven in verschillende vakken sinds 2016 aangemoedigd om samen te werken om een multidisciplinair onderwijsproject te ontwikkelen, maar er is geen duidelijke focus op evaluatie. De samenwerking die wordt aangemoedigd, heeft echter betrekking op verschillende vakken en er is geen beleid vastgesteld met betrekking tot de samenwerking van docenten in hetzelfde vak, met name wiskundedocenten. In Cyprus bestaat überhaupt geen vast beleid inzake samenwerking tussen docenten en scholen en het opzetten van netwerken voor het evalueren van leerlingen. Het is echter belangrijk op te merken dat alle geïnterviewde Cypriotische beleidsmakers bevestigen dat dit belangrijke elementen zijn om de kwaliteit van de leerkrachten te verbeteren.

Het is moeilijk om mogelijkheden te creëren voor samenwerking tussen docenten en het opzetten van een netwerk van scholen. Om de samenwerking doeltreffend te laten zijn, moet aan bepaalde voorwaarden worden voldaan. Bijvoorbeeld, geplande gemeenschappelijke planningstijd (Main, 2012), consistente vergadertijd (Saunders et al., 2009) en samenwerkingsbeleid en verantwoordingsystemen (George & Alexander, 2003). Hoewel samenwerking op zowel het niveau van de docenten als dat van de scholen door de beleidsmakers in alle vier de landen als belangrijk wordt erkend, worden de voorwaarden voor samenwerking op hoog niveau in geen enkel land gecreëerd.

In dit deel werd gekeken naar het beleid met betrekking tot de opleiding van leerkrachten vóór en tijdens de dienst, en de samenwerking tussen leerkrachten en scholen op het gebied van evaluatie. Op basis van het bovenstaande komen we tot de conclusie dat, hoewel er een aantal formele en informele mogelijkheden voor lerarenopleidingen op het gebied van assessment beschikbaar zijn, sommige leraren wellicht niet voldoende opgeleid zijn om hun leerlingen te beoordelen. Uit de analyse van de

gegevens is gebleken dat leraren in alle vier de landen worden aangemoedigd om van de geboden mogelijkheden gebruik te maken; er is echter geen sprake van een systematische opleiding om de kwaliteit van de evaluatiespraktijk (en met name de evaluatie in de wiskunde) te waarborgen.

3.3. Evaluatiemechanismen die betrekking hebben op het beleid in leerlingevaluatie en TPD

Steeds vaker wordt beleid geëvalueerd op basis van een aantal empirische bewezen evaluatiemechanismen: relevantie, impact, doeltreffendheid, efficiëntie, en duurzaamheid. Deze evaluatiemechanismen leveren geloofwaardige, betrouwbare en nuttige informatie op die kan worden gebruikt om de beleidsvorming te genereren en/of te verbeteren (Sanderson, 2002). Het derde aspect dat in dit verslag werd onderzocht, heeft betrekking op beleid dat relevant is voor mechanismen die door onderwijssystemen zijn ontwikkeld om het bestaande beleid inzake leerlingevaluatie en professionele ontwikkeling van docenten te kunnen evalueren. Bij de aanwezigheid van een evaluatiemechanisme in een land, is er daarna gekeken naar hoe vaak deze evaluatie wordt uitgevoerd en op hoeveel verschillende bronnen deze evaluatie wordt gebaseerd (d.w.z. meningen van verschillende belanghebbenden). Daarnaast werd de fase van de beleidsuitvoering waarin de evaluatie wordt uitgevoerd, wordt ook onderzocht (bv. aan het einde van de uitvoering, op regelmatige tijdstippen tijdens de uitvoering). Dit geeft weer wat de continuïteit van de evaluatie is en de mogelijkheden tot aanpassingen op basis van evaluatieresultaten. Bovendien hebben we gekeken naar de kwaliteit van de gebruikte processen om zo te verhelderen welke gegevens en bronnen kunnen worden gebruikt om beleidsmakers te helpen bij het nemen van beslissingen die het nationale beleid kunnen verbeteren. Ten slotte is in dit deel onderzocht hoe en voor wie de resultaten uit een evaluatie worden gebruikt. Zo kunnen we nagaan of het evaluatiemechanisme wordt gebruikt voor formatieve of summatieve doeleinden en of belanghebbenden toegang hebben tot informatie die hen kan helpen bij het aanbrengen van wijzigingen voor verbetering.

Op basis van de analyse van het beleid bestaan er in drie van de vier deelnemende landen (Cyprus, Griekenland en België) geen evaluatiemechanismen met betrekking tot het beleid inzake evaluatie en TPD. Opvallend is het principe van de onderwijsvrijheid in België en Nederland, want deze geeft scholen het recht om een schoolspecifieke visie op de kwaliteit van het onderwijs te

ontwikkelen binnen de grenzen die door de Vlaamse of Nederlandse regering zijn vastgesteld. Deze visie wordt uitgewerkt in een schoolspecifiek leerplan en een schoolspecifieke onderwijs- en evaluatiepraktijk. Het evaluaties- en evaluatiebeleid behoort dus tot de volledige autonomie van elke school. In België bestaan verder geen nationale evaluatiemechanismen. In dezelfde lijn, hebben ook de landen Cyprus en Griekenland, ondanks het aantal beleidsmaatregelen dat bepaalt hoe en wanneer de evaluatie plaatsvindt (in het geval van de summatieve evaluatie), geen gevestigde evaluatiemechanismen van deze beleidsmaatregelen, zelfs niet die verwijzen naar summatieve evaluatie. In beide landen zijn er bijvoorbeeld examens aan het eind van de twaalfde klas om leerlingen de mogelijkheid te geven om een plek op de plaatselijke universiteit te verwerven. Ondanks het grote belang van deze examens en hun impact op het leven van de leerlingen, bestaat er geen meta-evaluatiemechanisme van deze examens om de uitvoering ervan te controleren en te evalueren. Bovendien wordt van scholen verwacht dat ze een voorgesteld of vereist beleid implementeren, maar ze hebben geen verantwoordelijkheid om de implementatie ervan te controleren. Zoals vermeld in de interviews met de beleidsmakers in beide landen (d.w.z. drie beleidsmakers per land), kan het beleid in beide landen worden beëindigd, vervangen of gewijzigd, maar dit is meestal gebaseerd op herzien of nieuw onderwijsbeleid, politieke prioriteiten of onderwijstrends (zoals in het geval van formatieve evaluatie). Er vindt geen evaluatie plaats van de uitvoering van het beleid of de impact op het leren.

Het enige land van de vier deelnemende landen dat van de vier landen enig beleid heeft met betrekking tot evaluatiemechanismen is Nederland, hoewel niet alle beleidsmakers dit in hun interviews hebben vermeld. Op basis van de analyse van de beleidsdocumenten geeft de Nederlandse wet op basis van de analyse van de beleidsdocumenten de opdracht om nieuwe wetten vijf jaar na invoering te evalueren. Daarom wordt verwacht dat elke nieuwe wet met betrekking tot de leerlingevaluatie en/of de TPD van studenten een evaluatieclausule bevat. Dit staat ook voor afzonderlijke beleidsinitiatieven die bij ministerieel besluit worden geïntroduceerd, vaak halverwege de looptijd van het ministerieel besluit en aan het eind van de looptijd van het besluit. Hiervoor zal in de meeste gevallen een apart monitoringprogramma worden besteld, meestal uitgevoerd door een onafhankelijke onderzoeksinstelling. In een dergelijk programma wordt onderzocht of het beleidsinitiatief het gewenste effect heeft en of de beleidsregel in het algemeen effectief is. Daarnaast worden voorlopige en

eindevaluaties geschreven, die als basis dienen voor beëindiging, verlenging zoals die is, verlenging na wijziging, of die de basis kunnen vormen voor een wetswijziging. Het gesubsidieerde beleid wordt doorgaans het hele jaar door informeel gemonitord door de ambtenaar van het ministerie die belast is met het functioneren van de beleidsmaatregel. Een jaarlijks activiteitenverslag en financieel overzicht is een verplicht onderdeel van het subsidieproces. Als er een monitoringprogramma wordt opgezet, hangt het af van de methodiek hoe vaak gegevens worden verzameld, maar over het algemeen verwacht het ministerie één keer per jaar (voorlopige) resultaten. Bij het onderzoek naar de validiteit van de gebruikte evaluatie-instrumenten blijkt uit de geanalyseerde beleidsdocumenten dat ervaringen uit onafhankelijke wetenschappelijke studies en uit eerdere (gesubsidieerde) onderzoeksprojecten worden gebruikt voor het ontwerp en de implementatie van de evaluatiemechanismen. Wanneer een monitoringprogramma voor een specifieke beleidsmaatregel noodzakelijk is, worden aanbestedingen (d.w.z. academische en/of instellingen) betrokken en wordt van hen verwacht dat zij het programma volgens goede wetenschappelijke methoden uitvoeren. Daarnaast worden grotere monitoringprogramma's meestal begeleid door een commissie waarin wetenschappelijke, onderwijs- en beleidsdeskundigen zitting hebben.

De beleidsdocumenten in Nederland suggereren daarnaast dat het Ministerie van Onderwijs, Cultuur en Wetenschap voortdurend in dialoog is met de belanghebbenden. Belanghebbenden zijn onder meer lerarenorganisaties, de organisaties van de verschillende vakgebieden, de belangenorganisatie van scholen en schoolleiders (Vo-Raad), studentenorganisaties (LAKS) en ouders (Ouders & Onderwijs). Wie precies wordt geraadpleegd kan variëren en is afhankelijk van de specifieke beleidsmaatregel. De perspectieven van scholen en schoolleiders, docenten, leerlingen en ouders worden echter meestal wel meegenomen. Omdat het Nederlandse voortgezet onderwijs voor de meeste vakken eindigt met een centraal eindexamen, is het bovendien gebruikelijk om de effecten van een beleid te meten op de cijfers van de leerlingen die door de specifieke beleidsmaatregel zijn getroffen. Wanneer de cijfers van studenten niet als een geschikte maatstaf worden beschouwd, worden andere, van het beleid afhankelijke maatregelen gebruikt, zoals de motivatie, de beheersing en het welzijn van studenten. Ten slotte erkent het systeem, zoals vermeld in de interviews met beleidsmakers, dat publieke

en politieke acceptatie een voorwaarde is voor de effectieve uitvoering van een beleid. Zo wordt ook rekening gehouden met de vraag of een maatregel erg populair of impopulair is.

Uit het bovenstaande blijkt dat Nederland erkent dat evaluatie een waardevol inzicht kan bieden in de uitvoering van een beleid. De vastgestelde evaluatiemechanismen hebben echter alleen betrekking op beleid waarvoor scholen of leerkrachten verantwoordelijk worden gesteld. Aan de andere kant worden, zoals uit de analyse van de gegevens blijkt, de meeste evaluatie-gerelateerde beleidsmaatregelen als aanmoedigingen genoemd en krijgen scholen en docenten de autonomie om een basis te formuleren en uit te voeren op basis van hun eigen behoeften en oordeel, zonder dat zij daarvoor verantwoordelijk worden gesteld. Dit betekent dat de evaluatieclausule in deze gevallen van toepassing is.

Als we kijken naar het derde en laatste onderzochte aspect, komen we tot de conclusie dat geen enkel ander deelnemend land (d.w.z. Cyprus, Griekenland en België), behalve Nederland, mechanismen heeft opgezet om zijn beleid inzake evaluatie en TPD te evalueren. Dit wordt als problematisch gezien het feit dat de evaluatie van beleidshervormingen belangrijke inzichten kan geven in de wijze waarop het hervormingsbeleid wordt uitgevoerd en gegevens oplevert op basis waarvan een beleid kan worden voortgezet, beëindigd of verbeterd om zijn doel te bereiken. Het FORMAS-project tracht het gebruik van evaluatie voor formatieve doeleinden aan te moedigen en te ondersteunen; deze formatieve oriëntatie is echter niet evident op beleidsniveau. Hoewel beleidsmakers het belang en de waarde van formatieve evaluatie erkennen, lijken zij zich niet bezig te houden met een formatieve evaluatie van hun eigen beleid.

Conclusie

Het FORMAS-project probeert beleidsmakers aan te moedigen om het beleid rondom leerlingevaluatie te hervormen en mechanismen in te voeren voor de ondersteuning van docenten voor een effectieve uitvoering van formatief evalueren in de klas. Om dit te doen, is het huidige nationale beleid van de deelnemende landen aan dit project op het gebied van leerlingevaluatie geanalyseerd. De analyses die op basis van de beleidsdocumenten en de interviews met de beleidsmakers werden uitgevoerd, hielpen bij het identificeren van overeenkomsten en verschillen tussen de vier deelnemende landen.

Ten eerste werd vastgesteld dat in veel beleidsdocumenten in alle vier de landen geen expliciete verwijzing naar het doel van leerlingevaluatie voorkomt, wat erop wijst dat de verschillen tussen de twee doelstellingen van evalueren niet voldoende worden erkend. Dit wordt als problematisch beschouwd, aangezien in de literatuur wordt gesteld dat het niet mogelijk is om beide doelen met één enkel mechanisme te bereiken. Daarom is het belangrijk dat bij de invoering van beleidsmaatregelen met betrekking tot evalueren het doel van de evaluatie duidelijk wordt aangegeven. Verduidelijking van het doel van een evaluatie is noodzakelijk om de juiste procedures, methoden en instrumenten te kunnen gebruiken. Het is ook nodig om mogelijke misvattingen te voorkomen, zowel wat betreft de praktijk van de docenten als wat betreft de ondersteuning van de docenten. Ten slotte is verduidelijking van het doel belangrijk om alle belanghebbenden in staat te stellen zowel de doeltreffendheid van een beleidsmaatregel als de kwaliteit van de uitvoering ervan te evalueren.

Ondanks het feit dat de meeste onderzochte beleidsmaatregelen geen expliciete verwijzing bevatten naar het doel van de evaluatie, zijn er in alle vier de landen beleidsmaatregelen die erop wijzen dat het formatieve doel van de evaluatie zeer gewaardeerd wordt en erkend wordt als een middel om het leren van studenten te verbeteren en in te spelen op de gedifferentieerde leerbehoeften van studenten. Vooral de laatste jaren kunnen in alle vier de landen een aantal beleidlijnen met een formatieve oriëntatie worden geïdentificeerd. De meeste zijn echter slechts suggesties en het is aan de scholen of docenten om te beslissen of ze er gevolg aan zullen geven. Bovendien schetst dit beleid meestal wat de literatuur als een goede evaluatiepraktijk suggereert, maar er worden geen specifieke acties of stappen beschreven die een meer gedetailleerde begeleiding en ondersteuning kunnen bieden aan de beoogde gebruikers. Concreet betekent dit dat ondersteuning in de vorm van professionele ontwikkeling, advies, begeleiding of middelen alleen beschikbaar is op vrijwillige basis en meestal niet in een gestructureerd kader. Aan de andere kant is het beleid inzake summatieve evaluatie meestal veel specifieker en bevat het meestal een eisenclausule. Gezien het verband tussen summatieve evaluatie, verantwoording en “high-stakes” beslissingen, is het niet verrassend dat docenten meer nadruk leggen op het gebruik van evaluatie voor summatieve in plaats van formatieve doeleinden. Bovendien is de uitvoering van een formatieve evaluatie een grotere uitdaging, omdat het meer tijd en moeite kost om deze voor hen uit te voeren. Onderzoek suggereert namelijk dat leraren positief lijken te staan tegenover een evaluatie die

het leren bevordert (Brown, 2004; Sach, 2012), maar dat hun praktijk nog steeds resultaatgericht lijkt te zijn (Earl & Katz, 2000; Kahn, 2000; Lock & Munby, 2000; Christoforidou & Xirafidou, 2014). Om ervoor te zorgen dat docenten meer aandacht besteden aan formatieve evaluaties in de klas, is het niet voldoende dat beleidsmaatregelen het formatieve doel van evaluaties en de impact ervan op het leren van studenten erkennen, maar moet deze ook de ondersteuning op dit gebied benadrukt worden. Uit de analyse is gebleken dat er geen beleid aanwezig is dat vereist dat zij over de nodige kennis en vaardigheden op het gebied van evalueren beschikken. Bovendien worden de meeste opleidingsmogelijkheden op vrijwillige basis aangeboden. Zoals de literatuur bevestigt, melden docenten onvoldoende training in klassikale evaluatie, zowel tijdens hun lerarenvoorbereidingsprogramma's als tijdens hun bijscholing (DeLuca & Klinger, 2010). De uitdagingen van de beschikbaarheid en de toereikendheid van de middelen die een doeltreffende evaluatiepraktijk ondersteunen zijn dus belangrijke onderdelen die aandacht nodig hebben. Dit vraagt om de invoering van beleidsmaatregelen die professionele normen schetsen voor de evaluatiepraktijk en in het bijzonder de implementatie van formatieve evaluatie in de wiskunde. Daarnaast is het niet voldoende dat beleidsmaatregelen het belang van de formatieve evaluatiepraktijk erkennen, maar ze moeten ook manieren vinden om docenten verantwoordelijk te stellen voor de uitvoering ervan.

Een andere kwestie die aan de orde is gesteld, heeft te maken met de opleiding van docenten in het evalueren. Uit de analyse bleek dat er beleid nodig is dat ervoor zorgt dat alle leerkrachten voldoende zijn opgeleid door middel van passende interventies op het gebied van professionele ontwikkeling, zowel wat betreft de inhoud als de duur van de opleiding. Hoewel leerkrachten unieke sterke en zwakke punten hebben, krijgen zij meestal "one-size-fits-all" professionele ontwikkeling, zoals eenmalige workshops, die de kwaliteit van hun evaluatiepraktijk en het leren minder snel verbeteren.

Het FORMAS-project heeft tot doel een alomvattende dynamische aanpak van TPD te ontwikkelen en te testen om docenten te ondersteunen bij het uitvoeren van leerlingevaluaties voor formatieve doeleinden. Deze aanpak erkent dat de professionele behoeften van de docenten met betrekking tot evaluaties ten eerste moeten worden geïdentificeerd en daarom is er een uitgebreid kader voor het meten van de vaardigheden van leerkrachten ontwikkeld (zie deliverable 2.2). Op basis van dit kader is een instrument voor het meten van de vaardigheden van leerkrachten op het gebied van

evalueren (zie resultaat 3.3) ontworpen. Dit instrument zal het worden toegepast vóór de interventie voor professionele ontwikkeling om op deze manier de professionele behoeften van docenten met betrekking tot evaluatie te identificeren. Dit zal ons in staat stellen om een meer gericht programma voor professionele ontwikkeling te bieden dat aansluit bij de geïdentificeerde behoeften van elke groep docenten. Bovendien wordt verwacht dat effectieve professionele ontwikkeling zal leiden tot veranderingen in de praktijk van de docenten en het leren van studenten (Darling-Hammond, Hylar, & Gardner, 2017). Daarom zal het FORMAS-project een experimenteel onderzoek uitvoeren naar de impact van de dynamische benadering van TPD op de verbetering van de evaluatievaardigheden van docenten en op de bevordering van de cognitieve en metacognitieve leerresultaten van leerlingen. Op basis van de resultaten van het onderzoek zullen we in staat zijn om specifieke en evidence-based beleidsrichtlijnen en relevante ondersteunende mechanismen aan te reiken die kunnen helpen bij de effectieve implementatie van formatieve evaluatiepraktijken. Het project beoogt via deze studie ook het bewustzijn te ontwikkelen bij beleidsmakers voor het introduceren van een evidence-based en theory-driven aanpak in de vorming en implementatie van evaluatie-gerelateerd beleid.

Bovendien heeft de analyse van het beschikbare beleid in de vier deelnemende landen aan het licht gebracht dat het beschikbare beleid niet gericht is op specifieke onderwerpen of leeftijdsgroepen. Tijdens de analyse is gebleken dat bijna alle evaluatiebeleidsmaatregelen geen specifieke verwijzing naar het onderwerp wiskunde bevatten. Dit is een belangrijke bevinding omdat de huidige opvattingen over effectieve wiskundige instructie de complexiteit van wiskunde op waarde schatten (Boaler, 2008) en vereisen dat docenten de nodige vaardigheden hebben ontwikkeld om het conceptuele begrip van studenten en hun probleemoplossend en redeneervermogen te evalueren (Suurtamm e.a., 2010). Beleid dat zich specifiek richt op de evaluatie van de wiskundekennis en -vaardigheden van leerlingen wordt daarom noodzakelijk geacht.

Ten slotte bleek uit de analyse dat er in drie van de vier deelnemende landen (Cyprus, Griekenland en België) geen evaluatiemechanismen bestaan. In Nederland zijn evaluatiemechanismen aanwezig in het geval een onderwijswet of -beleid een verantwoordingselement heeft. Er werden echter geen evaluatiemechanismen met een formatieve functie in het beleid vastgesteld. Dit komt omdat de meeste beleidsmaatregelen die verwijzen naar een formatieve evaluatie van de leerlingen,

aanbevelingen zijn en het is aan de scholen en docenten om te beslissen of ze deze aanbevelingen al dan niet zullen opvolgen.

Monitoring- en evaluatiemechanismen worden als essentiële instrumenten beschouwd, omdat ze de besluitvorming beïnvloeden op basis van de lessen die men heeft geleerd. Op basis van de resultaten van de evaluatie worden beslissingen genomen om een interventie of beleid te verbeteren, te heroriënteren of stop te zetten. Vooral op het gebied van onderwijs is de evaluatie van interventies en/of beleid noodzakelijk omdat deze een impact hebben op leerlingen en hun leren. Daarom is de invoering van evaluatiemechanismen voor beleid dat verband houdt met de evaluatie in de onderwijspraktijk noodzakelijk. Onderzoek toont aan dat formatieve evaluatiepraktijken een positieve impact hebben op de prestaties van studenten (Hattie & Timperley, 2007). Om ervoor te zorgen dat formatieve evaluaties een concrete praktijk wordt bij (wiskunde)docenten in het secundair onderwijs, moet er echter ondersteuning op systeemniveau worden geboden. Het introduceren van nieuw beleid of het aanpassen van bestaand beleid op basis van de beschikbare gegevens en een theorie-gedreven aanpak zijn nodig om ervoor te zorgen dat het evalueren in de klas op de juiste manier wordt ondersteund en dat deze een positieve impact heeft op het leren van de leerlingen.

References

- Antoniou, P., & Kyriakides, L. (2013). A Dynamic Integrated Approach to Teacher Professional Development: Impact and Sustainability of the Effects on Improving Teacher Behavior and Student Outcomes. *Teaching and Teacher Education, 29*(1), 1-12.
- Berry, R. (2008). *Assessment for Learning*. Hong Kong: Hong.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice, 5*(1), 7-74.
- Boaler, J. (2008). *What's math got to do with it?* London: Penguin Press.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational researcher, 33*(8), 3-15.
- Broadfoot, P. (1992) Exploring the forgotten continent: a traveller's tale. *Scottish Educational Review, 26*(2), 88-96.
- Brookhart, S.M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice, 22*(4), 5–12.
- Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*(3), 301-318.
- Christoforidou, M., & Xirafidou, E. (2014). Using the dynamic model to identify stages of teacher skills in assessment. *Journal of Classroom Interaction, 49*(1), 12-25.
- Cohen, D.K., & Hill, H.C. (2001). *Learning policy*. New Haven, CT: Yale University Press.
- Creemers, B., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London and New York: Routledge.
- Creemers, B.P.M., & Kyriakides, L. (2015). Developing, testing and using theoretical models of educational effectiveness for promoting quality in education. *School Effectiveness and School Improvement, 26*(1), 102-119.
- Creemers, B. P., Kyriakides, L., & Antoniou, P. (2013). A dynamic approach to school improvement: Main features and impact. *School Leadership & Management, 33*(2), 114-132.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: a review of state policy evidence. *Education Policy Analysis Archives, 8*(1), <http://epaa.asu.edu/epaa/v8n1/>

- Darling-Hammond, L. (2006). Securing the right to learn: Policy and practice for powerful teaching and learning. *Educational Researcher*, 35(7), 13-24.
- Darling-Hammond, L., Hyler, M. E., Gardner, M. (2017). *Effective Teacher Professional Development*. Palo Alto, CA: Learning Policy Institute.
- DeLuca, C., & Klinger, D.A. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438.
- DES/WO (1988) *National Curriculum Task Group on Assessment and Testing—a report*. London: DES.
- Desimone, L. M. (2009). Improving impact studies of teacher's professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181-199.
- Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational evaluation and policy analysis*, 24(2), 81-112.
- Earl, L., & Katz, S. (2000). Changing classroom assessment: Teachers' struggles. In N. Bascia & A. Hargreaves (Eds.), *The sharp edge of educational change* (pp. 97-111). London: Routledge.
- George, P.S., & Alexander, W.M. (2003). *The exemplary middle school*. Wadsworth Publishing Company.
- Gipps, C. (1994). *Beyond testing*. London: Routledge Falmer.
- Guskey, T. R. (2003). What makes professional development effective? *Phi delta kappan*, 84(10), 748-750.
- Guskey, T. R., & Sparks, D (2004). Linking professional development to improvements in student learning. In E.M. Guyton & J.R. Dangel (Eds.), *Research Linking Teacher Preparation and Student Performance: Teacher Education Yearbook XII*. Lanham, MD: Rowman & Littlefield Education.
- Hanushek, E.A. (1986). The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature*, 24, 1141-1177.
- Hanushek, E.A. (1989). The impact of differential expenditures on student performance. *Educational Research*, 66(3), 397-409.

- Harlen, W., & James, M.J. (1997). Assessment and learning: differences and relationship between formative and summative assessment. *Assessment in Education*, 4(3), 365-380.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112.
- Hedges, L.V., Laine, R.D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes (An exchange: Part 1). *Educational Researcher*, 23(3), 5-14.
- Herman, J.L., Osmundson, E., Ayala, C., Schneider, S., & Timms, M. (2006). *The nature and impact of teachers' formative assessment practices. CSE Technical Report #703*. National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Khan, E. A. (2000). A case study of assessment in a grade 10 English course. *The Journal of Educational Research*, 93, 276-286.
- Kennedy, M. M. (2016). How does professional development improve teaching? *Review of educational research*, 86(4), 945-980.
- Kyriakides, L. (2005). Extending the comprehensive model of educational effectiveness by an empirical investigation. *School Effectiveness and School Improvement*, 16(2), 103–152.
- Kyriakides, L., & Campbell, R.J. (2003). Teacher Evaluation in Cyprus: Some conceptual and methodological issues arising from Teacher and School Effectiveness Research. *Journal of Personnel Evaluation in Education*, 17(1), 21-40.
- Kyriakides, L., & Creemers, B.P.M. (2008). Using a multidimensional approach to measure the impact of classroom level factors upon student achievement: a study testing the validity of the dynamic model. *School Effectiveness and School Improvement*, 19(2), 183-205.
- Kyriakides, L., Creemers, B.P.M., & Charalambous, E. (2018). *Equity and Quality Dimensions in Educational Effectiveness*. Dordrecht, the Netherlands: Springer.
- Kyriakides, L., Creemers, B.P.M., Antoniou, P., Demetriou, D., & Charalambous, C. (2015). The impact of school policy and stakeholders' actions on student learning: A longitudinal study. *Learning and Instruction*, 36, 113-124.

- Kyriakides, L., Demetriou, D., & Charalambous, C. (2006). Generating criteria for evaluating teachers through teacher effectiveness research. *Educational Research, 48*(1), 1-20.
- Kyriakides, L., Georgiou, M.P., Creemers, B.P.M., Panayiotou, A., & Reynolds, D. (2018). The impact of national educational policies on student achievement: a European study. *School Effectiveness and School Improvement, 29*(2), 171-203.
- Lock, C.L., & Munby, H. (2000). Changing assessment practices in the classroom: A study of one teacher's change. *The Alberta Journal of Educational Research, 46*(3), 267-279.
- Loucks-Horsley, S., Hewson, P.W, Love, N. & Stiles, K. (1998). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Corwin.
- Main, K. (2012). Effective middle school teacher teams: A ternary model of interdependency rather than a catch phrase. *Teachers and Teaching, 18*(1), 75-88.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Richter, D., Kunter, M., Klusmann, U., Lüdtke, O., & Baumert, J. (2014). Professional development across the teaching career: Teachers' uptake of formal and informal learning opportunities. In *Teachers' Professional Development* (pp. 97-121). Brill Sense.
- Roeber, E.D. (2003). *Assessment models for No Child Left Behind*. Denver, CO: Education Commission of the States.
- Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal, 52*(3), 475-514.
- Sach, E. (2012). Teachers and testing: an investigation into teachers' perceptions of formative assessment. *Educational Studies, 38*(2), 261-276.
- Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public administration, 80*(1), 1-22.
- Saunders, W.M., Goldenberg, C.N., & G. Gallimore, R. (2009, December). Increasing achievement by focusing grade-level teams on improving classroom learning: A prospective, quasi-experimental study of Title I schools. *American Educational Research Journal, 46*(4), 1006-1033.

- Scheerens, J. (2016). *Educational Effectiveness and Ineffectiveness: A Critical Review of the Knowledge Base*. Dordrecht, The Netherlands: Springer.
- Smith, M.S., & O'Day, J. (1991). Systemic school reform. In S. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233-268). Philadelphia: Falmer Press.
- Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' emerging assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy, and Practice*, 17(4), 399-417.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: teaching, learning and assessment in the classroom*. Buckingham, UK: Open University Press.
- Vangrieken, K., Dochy, F., Raes, E., & Kyndt, E. (2015). Teacher collaboration: A systematic review. *Educational research review*, 15, 17-40.
- William, D., Lee, C., Harrison, C., & Black, P. J. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles Policy and Practice*, 11(1), 49-65.