



Erasmus+

Mede gefinancierd door het
programma Erasmus+
van de Europese Unie



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



Effectieve evaluatiepraktijk: een uitgebreid kader voor het meten van de evaluatievaardigheden van leraren

Dit verslag presenteert een raamwerk voor het definiëren en meten van evaluatievaardigheden van leraren. Het raamwerk is beschreven voor het Erasmus+KA3 project genaamd: “Bevorderen van Formatief Evalueren: Van Theorie naar Beleid en Praktijk”. Dit raamwerk past bij de huidige percepties van effectief lesgeven en evalueren en maakt het mogelijk om de behoeften van leraren op dit gebied te identificeren. Het voorgestelde kader houdt rekening met het dynamische karakter van onderwijs en de vaardigheden van elke fase van de evaluatiepraktijk. Daarnaast zullen evaluatievaardigheden gedefinieerd worden en gemeten in relatie tot het vermogen van leraren om specifieke evaluatietechnieken in te zetten om verschillende leeruitkomsten op leerlingenniveau te kunnen meten.

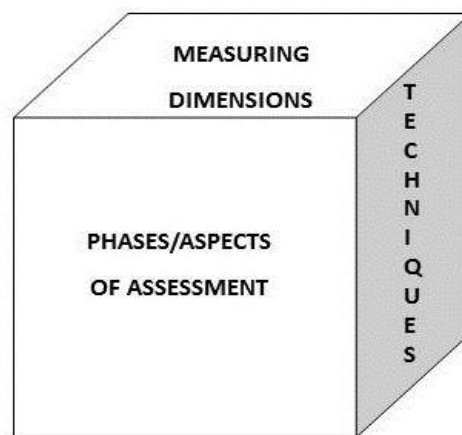
1. Introductie

In de afgelopen decennia is klassikale evaluatie een belangrijk onderdeel geweest van verschillende inspanningen op het gebied van onderwijsverbetering. Hierbij werd de nadruk gelegd op de rol van de evaluatiepraktijk bij het ondersteunen van het leren van leerlingen. Een aantal effectiviteitsstudies hebben evaluatie als een belangrijke effectiviteitsfactor aangemerkt (bijvoorbeeld Antoniou & Kyriakides, 2011; Hattie & Timperley, 2007), waardoor het verbeteren van de evaluatiepraktijk een noodzaak is geworden. Het doel van dit rapport is om een raamwerk te bieden dat ons in staat stelt om het evaluatiegedrag van leraren te onderzoeken en om specifieke vaardigheden te identificeren die betrokken zijn bij het beoordelen van het leerproces van leerlingen. Dit is belangrijk, omdat eerdere pogingen om te definiëren wat leraren zouden moeten weten en kunnen doen in relatie tot evaluatie onsystematisch zijn geweest (Brookhart, 2011; Mok, 2010). Bovendien zal dit kader deze evaluatievaardigheden evalueren aan de hand van de impact op de leerprestaties van leerlingen (d.w.z. wiskunde). Op deze manier kunnen we professionele standaarden identificeren die effectieve

evaluatiepraktijken beschrijven en vervolgens een interventieprogramma voor professionele ontwikkeling ontwerpen dat een positief effect kan hebben op het verbeteren van de onderwijspraktijk en de leerprestatie van leerlingen. Bovendien is het de bedoeling kritisch in te gaan op de theorie van formatief evalueren (Black & Wiliam, 2006) zonder echter voorbij te gaan aan het feit dat een groot percentage van de evaluatiepraktijk meer summatief georiënteerd is (Herman, Osmundson, Ayala, Schneider, & Timms, 2006). Om deze redenen worden er in dit raamwerk vaardigheden beschreven die gerelateerd zijn aan evaluatie voor zowel **summatieve als formatieve** doeleinden. Hierdoor wordt het mogelijk om evaluatiegedrag te onderzoeken en goede evaluatiepraktijken te identificeren, ongeacht de doeloriëntatie. Het voorgestelde kader maakt het dus mogelijk de evaluatievaardigheden van leerkrachten te meten wanneer leraren leerlingen evalueren voor summatieve doelen, formatieve doeleinden of beide.

2. Een kader voor het meten van de toetsingsvaardigheden van leraren

Het in dit rapport gepresenteerde kader is gebaseerd op eerdere onderzoeken naar de evaluatievaardigheden van leraren in verschillende Europese landen (zie Christoforidou & Xirafidou, 2014; Christoforidou, Kyriakides, Antoniou, & Creemers, 2014). Het voorgestelde kader (zie figuur 1) onderzoekt evaluatie op basis van drie hoofdaspecten.



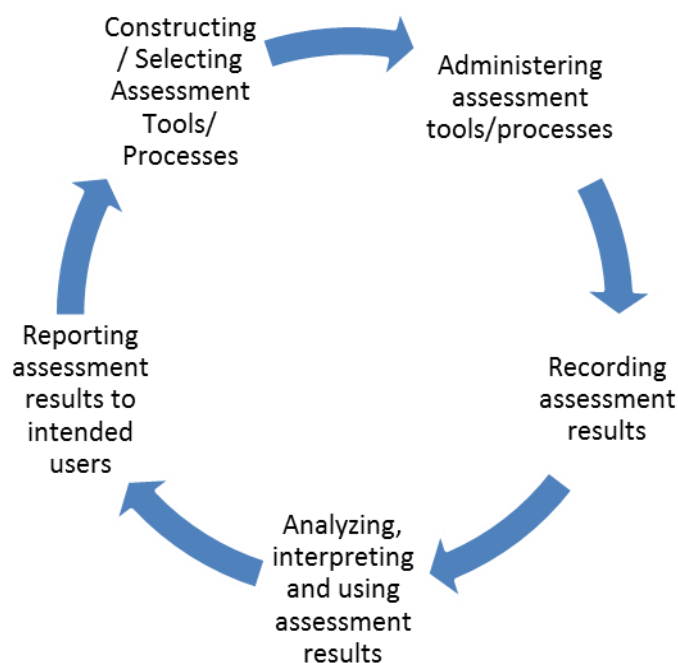
Figuur 1. Een raamwerk voor het meten van evaluatievaardigheden van leraren.

Ten eerste wordt rekening gehouden met het dynamische karakter van het onderwijs (Scheerens, 2016) en worden vaardigheden die verband houden met de hoofdfasen van het evaluatieproces (zie figuur 2) onderzocht (d.w.z. instrumenten / -processen construeren / selecteren, toetsingsinstrumenten / -processen beheren, vastleggen van resultaten, analyseren, interpreteren en gebruiken van resultaten,

rapporteren van resultaten aan beoogde gebruikers). Ten tweede worden evaluatievaardigheden gedefinieerd en gemeten in verhouding tot het vermogen van leraren om de belangrijkste evaluatietechnieken te gebruiken (d.w.z. schriftelijke beoordeling, mondelinge beoordeling, prestatiebeoordeling), evenals hun vaardigheden om niet alleen de leraar maar ook de zelf-, peer- en co-evaluatie in te zetten. Ten slotte wordt het meetkader voorgesteld van het *Dynamische Model van Educational Effectiveness* (Creemers & Kyriakides, 2008) aangenomen, waarmee we zowel kwantitatieve als kwalitatieve kenmerken van het evaluatieproces (d.w.z. frequentie, focus, fase, kwaliteit en differentiatie) kunnen onderzoeken. Een gedetailleerde presentatie van de drie belangrijkste aspecten wordt hieronder gepresenteerd.

a) Fasen van het evaluatieproces

Evaluatie wordt beschouwd als een integraal onderdeel van het onderwijs en wordt gedefinieerd als het systematische proces van het verzamelen van informatie over het leren van leerlingen (Shepard, 2000). Herkend als een proces, wordt evaluatie vaak in de literatuur gepresenteerd als een cyclus die is onderverdeeld in een aantal fasen (d.w.z. Birenbaum et al., 2009; Rea-Dickins, 2001; NCTM, 1995). Elke fase van het evaluatieproces kan worden gekenmerkt door de beslissingen en acties die binnen die fase plaatsvinden. Het voorgestelde raamwerk identificeert vijf hoofdfasen die op een alomvattende manier het proces van evaluatieontwerp en -praktijk beschrijven (zie figuur 2).



Figuur 2. De belangrijkste fasen van het toetsingsproces

Zoals weergegeven in figuur 2, worden vijf fasen geïdentificeerd. Deze fasen zijn gebaseerd op de aanname dat effectieve leerkrachten ervoor moeten zorgen dat: (i) geschikte toetsingsinstrumenten worden gebruikt om geldige en betrouwbare gegevens te verzamelen; (ii) passende procedures bij het beheer van deze instrumenten worden gevolgd; (iii) evaluatieresultaten op een efficiënte manier worden geregistreerd zonder daarbij belangrijke informatie te verliezen; (iv) evaluatieresultaten worden geanalyseerd, geïnterpreteerd en gebruikt op manieren die het leren van leerlingen kunnen bevorderen; en (v) evaluatieresultaten worden gerapporteerd aan alle beoogde gebruikers, inclusief ouders en leerlingen, om hen te helpen beslissingen te nemen over het verbeteren van de leerprestaties van leerlingen. De cyclus, zoals weergegeven in figuur 2, kan tot de conclusie leiden dat het eenvoudigweg een stapsgewijs model is dat door de leraar wordt uitgevoerd. Het is belangrijk op te merken dat de fasen niet worden beschouwd als een stapsgewijs model dat eenvoudigweg alleen door de leraar wordt uitgevoerd. Integendeel, het raamwerk is gebaseerd op het huidige denken in evaluatie dat het beschouwt als een lopend, iteratief, dynamisch proces dat zowel de leraar als de leerling in het proces betreft (Shepard, 2000; Gardner, Harlen, Hayward, Stobart, & Montgomery, 2010; Wiliam, Lee, Harrison, & Black, 2004). De literatuur belicht ook de dynamische relatie tussen de verschillende fasen van het evaluatieproces (Birenbaum, 2007; Black & Wiliam, 2009). Zonder het sequentiële karakter van de vijf fasen die betrokken zijn bij het proces van het ontwerp en de uitvoering van beoordelingen te verwaarlozen, beschouwt dit raamwerk alle fasen als onderling verbonden en onderling uitwisselbaar. De indeling van het evaluatieproces in bepaalde fasen wordt gedaan om ervoor te zorgen dat elk aspect van de evaluatiepraktijk in aanmerking wordt genomen bij het meten van evaluatievaardigheden. Een gedetailleerde beschrijving van elke fase wordt hieronder weergegeven.

i. evaluatiehulpmiddelen / -processen construeren / selecteren

Deze fase omvat vaardigheden die betrekking hebben op de planning en het ontwerpen van een evaluatie, maar ook op de selectie en / of constructie van de instrumenten en -processen. Ten eerste is het noodzakelijk voor leraren om te bedenken welk **doel** een toetsing dient om de geschikte procedures, methoden en instrumenten te kiezen. In deze fase wordt van leraren verwacht dat ze beslissen of ze een evaluatie plannen om summatieve of formatieve doelen te bereiken. Na het bepalen van het doel van een evaluatie, is het noodzakelijk om de leerdoelen te definiëren op basis waarvan een leerling wordt beoordeeld. Herman en collega's (2006) beschouwen leerdoelen als start-, eind- en recyclingpunten bij (1) de selectie en implementatie van kwaliteitstoetsingsinstrumenten, (2) de interpretatie en analyse van toetsingsresultaten, en (3) het gebruik van resultaten om informatieve feedback te geven en acties ondernemen die de leerresultaten van leerlingen verbeteren. Sterker nog, de rol van leerdoelen lijkt essentieel gedurende het gehele evaluatieproces. Doelen bieden een kader voor het interpreteren en reageren op gebeurtenissen die plaatsvinden (Yorke, 2003). Daarnaast beïnvloeden ze de prestaties door

aandacht te richten, inspanningen te mobiliseren, persistentie te verhogen en de ontwikkeling van leerstrategieën te stimuleren (Locke et al., 1981). Het delen van de leerdoelen en succescriteria die met de leerlingen (Black & Wiliam, 1998) is ook belangrijk, omdat het de mate van activiteit verhoogt met betrekking tot de rol van de leerling in het toetsingsproces. Ten slotte omvat deze fase de selectie of constructie van de nodige evaluatie-instrumenten en -processen. De term 'evaluatiertools' verwijst naar instrumenten, strategieën en processen die kunnen worden gebruikt om het leren van leerlingen te beoordelen, zoals observatieprotocollen, dialoog in de klas, bevringsstrategieën en zelftoetsingsrubrieken. Op deze manier omvat de term een bredere variëteit aan traditionele en niet-traditionele hulpmiddelen dan op dit moment in de praktijk. De evaluatiertools spelen een belangrijke rol bij het waarborgen van de kwaliteit en effectiviteit van toetsing en hebben een grote invloed op hoe en wat leerlingen leren (Boud, 1988; Brown, Bull & Pendlebury, 1997). Hoewel een aantal evaluatiertools beschikbaar zijn, kunnen deze niet door elkaar worden gebruikt. De reden hiervoor is dat bepaalde leerdoelen alleen te meten zijn met bepaalde toetsingstools. Daarnaast is de context ook een factor is waarmee rekening moet worden gehouden, omdat deze invloed heeft op hoe goed een gebruiker reageert op een specifieke tool (Stiggins, 1992). De vaardigheden die dus in deze fase van belang zijn, zijn (a) het kunnen maken van de juiste keuze voor een bepaalde evaluatietechniek op basis van het doel dat deze evaluatie dient (Brookhart, 2003; Gipps, 1994; Pellegrino, Chudowsky, & Glaser, 2001; Torrance & Pryor, 1998) en (b) het kunnen definiëren van de leerdoelen waartegen een leerling beoordeeld zal worden (Herman et al., 2006; Sadler, 1989). Bovendien omvat deze fase de vaardigheden die verband houden met de selectie en / of ontwikkeling van kwaliteitstoetsingsinstrumenten waarmee het doel en de doelen van de beoordeling worden bereikt (Green & Mantz, 2002; Shepard, 2000).

ii. Toetsingshulpmiddelen / -processen beheren

De tweede fase omvat vaardigheden die verband houden met de uitvoering van evaluatie. Terwijl externe evaluatie doorgaans meer gestandaardiseerd is in termen van timing, houding en ondersteuning van leerkrachten, berust veel evaluatie in de klas vooral op beslissingen van leerkrachten. Zoals Black en Wiliam (2006) betogen, steunt een effectieve evaluatie zwaar op de aanpassingen die leraren maken op basis van de resultaten van de evaluatie; deze aanpassingen variëren in zowel scope als timing. Leraren kunnen bijvoorbeeld de duur van een evaluatietaak differentiëren, instructies vereenvoudigen of zelfs de inhoud van een evaluatie differentiëren op basis van de behoeften van hun leerlingen. De vaardigheden die in deze fase van belang zijn, hebben betrekking op beslissingen die genomen moeten worden over de timing van een evaluatie, de koppeling van de evaluatie aan instructie, de verscheidenheid aan gebruikte technieken en de rol van leraren tijdens de evaluatie-administratie (Anderson, 2003; Black & Wiliam, 1998; Shepard, 2007).

iii. Evaluatieresultaten registreren

De derde fase verwijst naar vaardigheden die verband houden met de documentatie van resultaten die zijn afgeleid van het evaluatieproces. Het documenteren van informatie is noodzakelijk om informatie effectief te kunnen gebruiken om het leren en onderwijzen te informeren. Het documenteren van resultaten maakt deze zichtbaar en deelbaar en vergroot zo de mogelijkheid dat ze worden gebruikt om lesgeven en leren te informeren (Rinaldi, 2006). Documentatie maakt het mogelijk bewijs te leveren dat de prestaties beschikbaar zijn voor toekomstig gebruik, interpretatie en revisie en helpt ook bij het identificeren van lacunes in het leren van leerlingen (Goldhaber & Smith, 2002). Documentatie wordt ook gezien als een uitstekend hulpmiddel voor het communiceren van resultaten aan beoogde gebruikers (Goldhaber & Smith, 2002) zoals vereist in de vijfde fase van het evaluatieproces. Helaas wordt een enorm deel van de dagelijkse evaluatie in feite nooit gebruikt, omdat deze niet gedocumenteerd is (Schmoker, 2006). Dit kan deels worden toegeschreven aan het feit dat leraren een beperkt begrip hebben van het doel, het belang, het proces en het effectieve gebruik van documentatie (Kroeger & Cardy, 2006). Effectieve documentatie vereist het bijhouden van regelmatig bijgewerkte records van de voortgang van leerlingen, het vastleggen van resultaten op manieren die kunnen worden gebruikt om de behoeften van leerlingen te identificeren en leerlingen bij te staan in het bijhouden van gegevens (Harlen et al., 1992; Stiggins & Chappuis, 2005).

iv. Resultaten analyseren, interpreteren en gebruiken

De vierde fase verwijst naar vaardigheden die verband houden met de analyse, interpretatie en gebruik van resultaten. Het belang van deze fase is te vinden in het feit dat het geen zin heeft informatie te verzamelen tenzij er op kan worden gereageerd. Leraren worden beschouwd als de primaire gebruikers van de verzamelde informatie. Nadat de leraar resultaten heeft vastgelegd, moet hij / zij beslissingen nemen over hoe deze informatie zal worden gebruikt. Leraren gebruiken de evaluatieresultaten optimaal om responsieve veranderingen in instructie en leren te bewerkstelligen (Popham, 2006). Deze veranderingen moeten vroeg genoeg zijn in het besluitvormingsproces om het leren van leerlingen daadwerkelijk te beïnvloeden (Stiggins & Chappuis, 2008).

v. Resultaten rapporteren aan beoogde gebruikers

De vijfde en laatste fase verwijst naar vaardigheden die verband houden met de rapportage van resultaten aan beoogde gebruikers. De communicatie van evaluatieresultaten overbruggt de kloof tussen de geregistreerde evaluatiegegevens, hun analyse en interpretatie en hun gebruik door de betrokken deelnemers. Om de bedoelde gebruikers daadwerkelijk te laten reageren op informatie, moeten ze inderdaad eerst op de hoogte worden gesteld van dergelijke informatie. Rapportageprocedures leveren

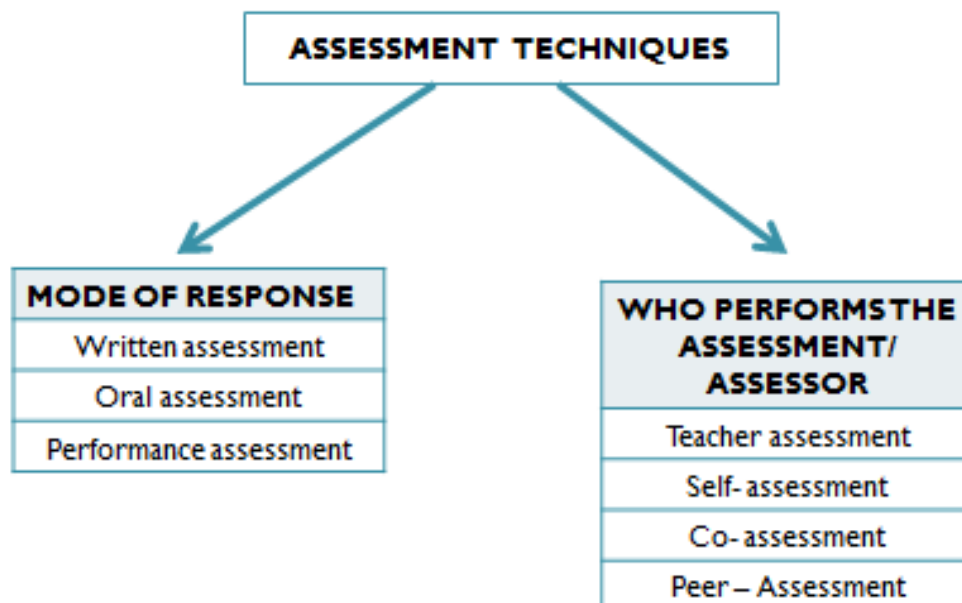
toetsingsresultaten in de handen van de verschillende beoogde gebruikers van de informatie op een tijdige en begrijpelijke manier (Roerber, 2003) en verbeteren de continuïteit en kwaliteit van de leerervaring van leerlingen (Berry, 2008). Deze rapportage biedt alle beoogde gebruikers (vooral ook leerlingen) de kans om hun strategieën effectiever in te zetten om hun leerprestatie te verbeteren. Het proces van het communiceren of rapporteren van resultaten brengt twee basisbeslissingen met zich mee: (1) wat is het doel dat de rapportage dient en (2) wat zijn de beste rapportagemethoden of hulpmiddelen om dit doel te bereiken (Guskey & Bailey, 2001; & James, 1997). Verschillende methoden kunnen worden gebruikt om de leervorderingen van leerlingen te rapporteren. De geselecteerde methode(s) moeten in overeenstemming zijn met het doel dat de beoordeling wil dienen en moeten op de juiste manier worden gebruikt om dit doel te bereiken. Stiggins (2004) suggereert bovendien dat effectieve communicatie van resultaten plaatsvindt wanneer iedereen de betekenis begrijpt van het prestatiedoel en de symbolen die worden gebruikt om informatie over te brengen, wanneer de informatie die ten grondslag ligt aan de communicatie juist is en ten slotte wanneer de communicatie is toegesneden op het beoogde publiek op het gebied van timing, detail en formaat. Vaardigheden die in deze fase zijn opgenomen, hebben betrekking op beslissingen over het doel van rapportage, het publiek van rapportage, de instrumenten die worden gebruikt om te rapporteren, evenals de kwaliteit van de communicatie tussen leraren met beoogde gebruikers.

b) Evaluatietechnieken

De term 'Evaluatietechnieken' verwijst naar de methoden die worden gebruikt om het leren van leerlingen te beoordelen. Evaluatiemethoden gaan over de instrumenten, strategieën en processen die kunnen worden gebruikt om het leren van leerlingen te beoordelen (bijvoorbeeld een schriftelijke test); "Evaluatietechnieken" is een specifiek begrip en verwijst naar het type evaluatiemethode dat kan worden toegepast (bijvoorbeeld geschreven evaluatie). Daarom wordt verwacht dat leraren eerst beslissen welke methode het meest geschikt is om te worden gebruikt (bijvoorbeeld mondelinge evaluatie) en vervolgens beslissen over het specifieke hulpmiddel dat moet worden ingezet (bijvoorbeeld een presentatie, mondelinge vraag, enz.). Bij het inzetten van evaluatietechnieken moet er rekening gehouden worden met twee belangrijke beslissingen die van invloed zijn op de selectie van evaluatietechnieken: a) de wijze van respons en b) degene die de evaluatie uitvoert. Deze combinatie verkrijgt de verzamelde informatie, omdat de evaluatietechnieken verschillende manieren van reageren vereisen en daarnaast ook door verschillende gebruikers ingezet kan worden (in de vormen van zelf-, peer- of co-assessment).

Op deze manier presenteert het kader evaluatie als een proces dat betrokkenheid van studenten verhoogt (Black & Wiliam, 2009; Harris & Brown, 2018; James & Pedder, 2006; Panadero, Jonsson, & Botella, 2017). Evaluatiemethoden spelen een belangrijke rol in het garanderen van de kwaliteit en de

effectiviteit van evaluatie, vermits ze doorgaans een invloed hebben op hoe en wat studenten leren. Het huidige denken rond evaluatie erkent dat een variatie van evaluatiemethoden gebruikt zou moeten worden. Leren is immers multidimensioneel en kan niet gepast gemeten worden door slechts een enkele methodiek (Brookhart, 2003; Gipps, 1994). Vertrouwen op slechts één methode zal daardoor slechts een stuk van de leerprestaties van leerlingen reflecteren. Een leraar zou bijvoorbeeld leerlingen de kans kunnen bieden om zowel mondeling als schriftelijk te tonen wat ze hebben geleerd (bijvoorbeeld door een mondelinge vraag en een schriftelijke oefening). De combinatie van deze twee evaluatievormen leiden tot betekenisvollere, meer valide en meer betrouwbare inzichten in het denken van leerlingen (Moss, 2003; Shepard, 2001). Daarom worden leraren aangemoedigd om een variatie aan evaluatievormen te gebruiken opdat leerlingen meerdere kansen hebben om te tonen wat ze weten en kunnen. De keuze voor een evaluatiemethode hangt af van het leerdoel dat geëvalueerd wordt, vermits leerlingprestaties voor bepaalde leerdoelen beter gemeten kan worden door specifieke methodieken (Stiggins, 1992). Het evalueren van mondelinge vaardigheden vereist bijvoorbeeld mondelinge oefeningen in plaats van schriftelijke toetsing. Daarbij komend laat een variatie in evaluatievormen studenten toe om verschillende types van leren te demonstreren. Dit is vooral zo in het kader van wiskunde omdat huidige inzichten in effectief wiskundeonderwijs de complexiteit van wiskunde onderschrijft (Boaler, 2008), wat van leraren vereist dat ze bekwaam zijn om verschillende evaluatievormen te gebruiken om zowel hun conceptuele begrip, hun probleemoplossend vermogen en hun redeneervaardigheden te evalueren (Suurtamm et al., 2010). Zoals hierboven beschreven, kijkt het raamwerk naar evaluatiemethoden door twee belangrijke beslissingen inzake de selectie van evaluatievormen in overweging te nemen: a) de manier van respons b) wie de performer is in de evaluatie (zie Figuur 3).



Figuur 3. Evaluatiemethoden

Wanneer we evaluatiemethoden categoriseren op basis van het type van respons kunnen we drie basistypes onderscheiden: *i) geschreven evaluatie*, *ii) mondelinge evaluatie*, en *iii) toepassingsgerichte evaluatie*. ***Geschreven evaluatie*** duidt op elke evaluatietaak die vereist dat leerlingen schriftelijk antwoorden. Dit type evaluatie verwijst doorgaans naar het gebruik van schriftelijke toetsen. Toch verwijst deze evaluatievorm ook naar quizzes, geschreven taken, schriftelijke oefeningen, rapporten en projecten. ***Mondelinge evaluatie*** duidt op elke evaluatietaak die een mondelinge respons vereist. Een voorbeeld is de leraar die mondeling vragen stelt of presentaties van leerlingen. ***Toepassingsgerichte evaluatie*** verwijst naar taken die vereisen dat leerlingen een product of antwoord creëren, of een specifieke set van taken uitvoeren om hun kennis en vaardigheden te demonstreren. Toepassingsgerichte evaluaties draaien rond een concrete product en/of prestatie die als bewijsvoering dient voor leren (bijvoorbeeld een 3D figuur van een specifiek volume creëren of een model bouwen). Toepassingsgerichte evaluatie is meteen gerelateerd aan observatie, gezien de beoordelaar verwacht wordt om de prestatie of het product te observeren om het leren te evalueren (Stiggins et al., 2006). Zowel incidentele als geplande observatie worden als noodzakelijk gezien bij het evalueren van het leren van leerlingen, vermits het een rijke en diverse bron van evidentie vormt om leeruitkomsten te evalueren.

Evaluatiemethoden worden ook onderverdeeld door middel van diegene die de rol van evaluator in het evaluatieproces inneemt. Op basis van dit criterium komen vier categorieën naar voren: *i) leraar evaluatie*, *ii) zelfevaluatie*, *ii) peer evaluatie*, and *iv) co-evaluatie*. ***Leraar evaluatie*** is de meest voorkomende vorm van evaluatie en duidt op de evaluaties waarin de leraar de verantwoordelijke is om het leren van leerlingen te evalueren. Het tweede type, zelfevaluatie, verlegt de rol van de beoordelaar van de leraar naar de leerling zelf. De argumentatie hiervoor is dat leerlingen geactiveerd worden als reflectief in hun leren (William & Thompson, 2007). Dit raamwerk gebruikt de invalshoek op zelfevaluatie als een soort instructiestrategie eerder dan een zelfregulerend proces (Panadero & Alonso-Tapia, 2013). Dit wil echter niet zeggen dat de impact van zelfevaluatie op de leervaardigheden en metacognitieve capaciteiten van leerlingen genegeerd wordt. Desalniettemin, gezien dit raamwerk leraarvaardigheden voor evalueren in kaart wil brengen wordt vooral gefocust op de rol van leraren in het promoten en gebruiken van zelfevaluatie en minder op de vaardigheden van leerlingen om dit ook effectief te doen. Hetzelfde gaat op voor peerevaluatie, het derde type dat onderscheiden wordt. ***Peerevaluatie*** verwijst naar evaluaties waarin peers elkaar evalueren. Opnieuw ligt de nadruk hier op de leraarvaardigheden in het promoten en implementeren van peerevaluatie in de klas. Ten slotte duidt het vierde type op een gezamenlijke methode van evalueren en kan elke combinatie van zelfevaluatie, peerevaluatie en evaluatie door de leraar hieronder vallen. In ***co-evaluatie*** is er een gedeeld doel om overeenstemming te vinden in de evaluatie (Dochy et al., 1999, p. 42).

Belangrijk om op te merken is dat, hoewel evaluatiemethoden onderscheiden kunnen worden op basis van het type leerlingrespons en diegene die de beoordelaar is, de twee categorieën noch afhankelijk zijn

van elkaar, noch elkaar uitsluiten. Dit wil zeggen dat elke evaluatievorm gebaseerd op het type antwoord van leerlingen kan uitgevoerd worden door verschillende of een combinatie van beoordelaars. Een toepassingsgerichte evaluatie kan bijvoorbeeld zowel door een leraar beoordeeld worden, als onderwerp zijn van een zelf-, peer-, of co-evaluatie. De nadruk in alle combinaties ligt op de vaardigheden van de leraar om elk van deze verschillende evaluatievormen te ontwerpen, uitvoeren, analyseren en rapporteren.

c) Meetdimensies

De dimensies die we onderscheiden om leraarvaardigheden te meten in evaluatie vloeien voort uit methodologische en theoretische ontwikkelingen in het expertisegebied van onderzoek naar onderwijseffectiviteit. Vroege studies in dit veld toonden aan dat kwantitatieve kenmerken van leraar evaluaties geassocieerd zijn met leerlingprestaties (bijvoorbeeld Scheerens & Bosker, 1997; Teddlie & Reynolds, 2000). Toch toonden recente studies aan dat ook kwalitatieve kenmerken van leraarevaluatie in rekening gebracht moeten worden (e.g., Heck & Moriyama, 2010; Kyriakides, 2005), vermits acties van leraren in de thematiek van evaluatie begrepen kunnen worden vanuit verschillende perspectieven en niet alleen door de mate waarin deze acties voorkomen in het evalueren van leerlingen.

De kwaliteit van evaluatie vereist niet louter een toename in frequentie, maar ook de nood om leerkanalen te creëren. Daarom worden vijf dimensies meegenomen die voortvloeien uit het dynamisch model van onderwijseffectiviteit (Creemers & Kyriakides, 2008). Concreet werde, de volgende vijf dimensies gebruikt om het functioneren binnen elk evaluatiekenmerk in te schatten: *(a) frequentie*, *(b) focus*, *(c) tijdstip*, *(d) kwaliteit* and *(e) differentiatie*. Deze dimensies helpen ons om het functioneren van effectieve leraren beter te beschrijven. Frequentie is een kwantitatieve manier om het functioneren te beschrijven, terwijl de andere vier dimensies kwalitatieve aspecten onderzoeken. Een korte beschrijving van de vijf dimensies van evalueren wordt in wat volgt gegeven. Het belang om elke dimensie meet e nemen word took geïllustreerd door te verklaren hoe de factor in het dynamisch model gemeten wordt.

Frequentie wordt gemeten door het aantal evaluatietaken van leraren in rekening te brengen, net als door de mate waarin evaluatie plaatsvindt. Deze meetdimensie helpt ons om het belang dat leraren aan evaluatie hechten in kaart te brengen. De andere dimensies bekijken kwalitatieve kenmerken van evaluatiepraktijken.

Focus wordt gemeten door te kijken naar de capaciteiten van leraren om een verscheidenheid in evaluatievormen te gebruiken in plaats van zich toe te spitsen op slechts één methodiek (Rao, Collins, & DiCarlo, 2002). Het is ook belangrijk om te onderzoeken of leraren de informatie die ze verzamelen voor meerdere doeleinden gebruiken (bijvoorbeeld de noden van leerlingen identificeren, aan

zelfevaluatie doen, lange termijn planningen aanpassen of evaluatietaken als een startpunt voor lesgeven gebruiken) (Black & Wiliam, 1998). Er zou een evenwicht moeten bestaan tussen heel specifieke en generieke taken. Als een enkel doel nagestreefd wordt, is de kans op success hoog, maar kan het effect echter slechts klein zijn doordat andere doelen niet bereikt worden en synergieën tussen afwezig blijven (Scheerens, 2013; Slater & Teddlie, 1992). Aan de andere kant, het nastreven van meerdere doelen houdt het gevaar in dat specifieke doelen niet op een manier geaddresserd worden dat ze succesvol geïmplementeerd kunnen worden. Vooral in het geval van evaluatie is dit belangrijk, gezien de twee hoofddoelen van evaluatie (summatieve en formatieve evaluatie) op andere waarden bouwen en dus tot verschillende praktische implementaties leiden.

De dimensie *tijdstip* wordt gemeten door te onderzoeken op welke tijdstippen de evaluatietaken plaatsvinden (bijvoorbeeld bij het begin, tijdens of op het einde van lessen of lessenreeksen) en welke tijdsperiode bestaat tussen het verzamelen van informatie, komen tot resultaten, deze te rapporteren naar leerlingen en ouders en deze te interpreteren en te gebruiken voor het plannen van lessen. We weten uit ander onderzoek dat activiteiten die bijdragen aan effectiviteit plaats zouden moeten vinden overheen een lange periode om te verzekeren dat ze een continue directe of indirecte impact hebben op het leren van leerlingen (Creemers, 1994; Slater & Teddlie, 1992). De verwachting is ook dat continuïteit bereikt wordt wanneer leraren flexibel zijn in het herdefiniëren van het eigen lesgeven en het de evaluatieactiviteiten aanpast door inzichten op basis van zelfevaluatie (Kyriakides & Campbell, 2004; Visscher & Coe, 2002).

Kwaliteit kan op twee verschillende manieren nagegaan worden. De eerste heft te maken met de kenmerken van de evaluatie, zoals ze bediscussieerd worden in de literatuur. Dus, kwaliteit wordt gemeten door naar de kenmerken van evaluatie-instrumenten te kijken, zoals de verschillende vormen van validiteit, de interne en externe betrouwbaarheid, de praktische haalbaarheid en de mate waarin de instrumenten de lesinhouden representeren (Cronbach, 1990). Het type feedback dat leraren geven aan leerlingen en de manier waarop leerlingen deze feedback gebruiken wordt ook onderzocht. Ten tweede, de impact van een evaluatie op leerlingprestaties wordt in rekening genomen. Zo is de impact van formatieve evaluatie op het leren van leerlingen breed gedocumenteerd (Black & Wiliam, 1998; Hattie & Timperley, 2007; Wiliam, Lee, Harrison, & Black, 2004). Daarom wordt van leraren verwacht dat zij meer nadruk leggen op formatieve doelen van evaluatie dan op summatieve doelen (Christoforidou, Kyriakides, Antoniou, & Creemers, 2014).

Ten slotte wordt *differentiatie* onderzocht wanneer het gaat om verschillende methoden om noden van leerlingen te meten en om verschillende manieren om feedback te geven aan verschillende groepen van leerlingen op basis van hun achtergrond en persoonlijke kenmerken. Het belang om differentiatie als een aparte dimensie te beschouwen komt voort uit het feit dat leerlingen van elke leeftijd en in elke cultuur van elkaar verschillen in cognitieve en niet-cognitieve vaardigheden, in generieke en specifieke

kennis, in hun interesses en motieven, in hun socio-economische achtergrond, en in persoonlijke denken en werkstijlen tijdens het leren (Dowson & McInerney, 2003). Onderzoekers in onderwijseffectiviteit hebben aangetoond dat deze verschillen gerelateerd zijn aan verschillen in leerwinst (Muijs et al., 2005).

Kenmerken van effectiviteit als multidimensionele constructen beschouwen leidt tot een beter begrip van wat leraren en scholen effectiever maakt, maar zou ook kunnen helpen in het ontwikkelen van heel specifieke strategieën om onderwijspraktijken te verbeteren (Kyriakides & Creemers, 2008). Dus, het gebruik van deze meetdimensies zal ons later toelaten om strategieën te ontwikkelen om evaluatiepraktijken te verbeteren, vermits de feedback aan leraren zich niet enkel op kwantitatieve maar ook op kwalitatieve kenmerken van hun evaluatiepraktijken zal richten.

1. Slotbemerkingen

Het erkennen van de nood aan een raamwerk voor het definiëren en identificeren van klasmanagement competenties heeft geleid tot het ontwikkelen van dit raamwerk. Het voorgestelde raamwerk benadert onderwijs met aandacht voor de dynamieken die spelen en onderzoekt vaardigheden die geassocieerd worden met elke fase van het evaluatieproces. Bijkomend zijn deze vaardigheden gedefinieerd en gemeten in relatie tot de capaciteit van leraren om specifieke evaluatietechnieken te gebruiken en op deze manier verschillende leeruitkomsten in het leergebied wiskunde te meten. Traditionele en alternatieve evaluatiemethoden die zowel gebaseerd zijn op de manier van antwoorden als op wie de evaluatie uitvoert zijn in overweging genomen gezien de literatuur het belang benadrukt van een combinatie van evaluatiemethoden om het leren van leerlingen te meten (Suurtamm et al., 2010). Meer nog, zowel kwalitatieve als kwantitatieve kenmerken van het evaluatieproces zijn mee in rekening genomen. Elk van de vijf evaluatiefases is gedefinieerd op basis van de kennis en vaardigheden die aanwezig zijn overheen de vijf meetdimensies en in relatie tot de meest voorkomende evaluatietechnieken.

Op basis van de literatuur is de basisassumptie dat leraarvaardigheden in evaluatie onderzocht moeten worden nog voor enige poging tot verbetering van evaluatiepraktijken kan plaatsvinden. Het raamwerk zal gebruikt worden om een onderzoeksinstrument (bijvoorbeeld, een vragenlijst) samen te stellen dat gebruikt kan worden om de vaardigheden van leraren met betrekking tot evaluatie te meten. De validatiestudie van dit instrument die in mei 2019 zal plaatsvinden in alle deelnemende landen zal ons toelaten om het raamwerk aan te passen en te verbeteren. Gezien voorgaand onderzoek erin geslaagd is om verschillende ontwikkelingsfasen van evaluatiegedrag te identificeren wordt het instrument ook geacht om specifieke noden van leraren in evaluatie bloot te leggen opdat gepaste acties ondernomen kunnen worden (Creemers, Kyriakides, & Antoniou, 2013).

Referenties

- Anderson, L. (2003). *Classroom assessment: Enhancing the quality of teacher decision making*. Mahwah, NJ: Erlbaum.
- Antoniou, P., & Kyriakides, L. (2011). The impact of a dynamic approach to professional development on teacher instruction and student learning: Results from an experimental study. *School Effectiveness and School Improvement*, 22(3), 291–311.
- Berry, R. (2008). *Assessment for Learning*. Hong Kong: Hong Kong University Press.
- Birenbaum, M. (2007). Assessment and instruction preferences and their relationship with test anxiety and learning strategies. *Higher Education*, 53(6), 749–768.
- Birenbaum, M., Kimron, H., Shilton, H., & Shahaf-Barzilay, R. (2009). Cycles of inquiry: Formative assessment in service of learning in classrooms and in school-based professional communities., *Studies in Educational Evaluation*, 35(4), 130–149.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7-74.
- Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 81-100). London: Sage
- Black, P., & Wiliam, D. (2009). Developing a theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.
- Boaler, J. (2008). *What's math got to do with it?* London: Penguin Press.
- Boud, D. (Ed.). (1988). *Developing Student Autonomy in Learning*, London: Kogan Page, Second Edition.
- Brookhart, S.M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5–12.
- Brookhart, S.M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3–12.
- Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing student learning in higher education*. London: Routledge.
- Christoforidou, M., Kyriakides, L., Antoniou, P., & Creemers. B.P.M. (2014). Searching for stages of teacher skills in assessment. *Studies in Educational Evaluation*, 40, 1-11.

- 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.
- Creemers, B. P. (1994). *The effective classroom*. Cassell.
- 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., & Antoniou, P. (2013). *Teacher professional development for improving quality of teaching*. Dordrecht, The Netherlands: Springer.
- Cronbach, L.J. (1990). *Essentials of psychological testing* (3rd ed.). New York: Harper & Row.
- Dochy, F.J.R.C., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher education*, 24(3), 331-350.
- Dowson, M., & McInerney, D.M. (2003). What do students say about their motivational goals?: Towards a more complex and dynamic perspective on student motivation. *Contemporary educational psychology*, 28(1), 91-113.
- Gardner, J., Harlen, W., Hayward, L., Stobart, G., & Montgomery, M. (2010). *Developing teacher assessment*. Maidenhead: Open University Press.
- Gipps, C. (1994). *Beyond testing*. London: Routledge Falmer.
- Goldhaber, J., & Smith, D. (2002). The development of documentation strategies to support Teacher reflection, inquiry, and collaboration. In V. Fu, A. Stremmel, & L. Hill (Eds.), *Teaching and learning: Collaborative exploration of the Reggio Emilia approach*, (pp. 147-160). Columbus, OH: Merrill/Prentice-Hall.
- Green, S.K., & Mantz, M. (2002, April). Classroom assessment practices: Examining impact on student learning. *Paper presented at the Annual Meeting of the American Educational Research Association*. New Orleans, LA.
- Guskey, T.R. & Bailey, J.M. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks, CA: Corwin Press, Inc.
- Harlen, W., & James, M.J. (1997). Assessment and learning: differences and relationship between formative and summative assessment. *Assessment in Education*, 4(3), 365-380.
- Harlen, W., Gipps, C., Broadfoot, P., & Nutall, D. (1992). Assessment and the improvement of education. *The Curriculum Journal*, 3(3), 215-230.
- Harris, L. R., & Brown, G. T. (2018). *Using Self-Assessment to Improve Student Learning*. Routledge.

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Heck, R. H., & Moriyama, K. (2010). Examining relationships among elementary schools' contexts, leadership, instructional practices, and added-year outcomes: A regression discontinuity approach. *School Effectiveness and School Improvement*, 21(4), 377–408.
- Herman, J.L., Osmundson, E., Ayala, C., Schneider, S., & Timms, M. (2006). The Nature and Impact of Teachers' Formative Assessment Practices. National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation Technical Report 703, University of California, Los Angeles, CA <http://www.cse.ucla.edu/products/reports/R703.pdf>.
- James, M., & Pedder, D. (2006). Professional learning as a condition for assessment for learning. In J. Gardner (Ed.), *Assessment and learning* (pp. 27–44). London: Sage.
- Kroeger, J., & Cardy, T. (2006). Documentation: A hard to reach place. *Early Childhood Education Journal*, 33(6), 389–398.
- 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. (2004). School self-evaluation and school improvement: A critique of values and procedures. *Studies in educational evaluation*, 30(1), 23-36.
- 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.
- Locke, E.A., Shaw, K.N., Saari, L.M., & Latham, G.P. (1981). Goal setting and task performance: 1969-1980. *Psychological Bulletin*, 90(1), 125-152.
- Mok, M.M.C. (2010). *Self-directed Learning Oriented Assessment: Assessment that Informs Learning & Empowers The Learner*. Hong Kong: Pace Publications Ltd.
- Moss, P.A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4), 13-25.
- Muijs, D., Campbell, J., Kyriakides, L., & Robinson, W. (2005). Making the case for differentiated teacher effectiveness: An overview of research in four key areas. *School Effectiveness and School Improvement*, 16(1), 51-70.
- National Council of Teachers of Mathematics (NCTM). (1995). *Assessment standards for school mathematics*. Reston, VA: Author.

- Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology, 11*(2), 551-576.
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review, 22*, 74-98.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Popham, W.J. (2006). All about accountability: Those [fill-in-the-blank] tests! *Educational Leadership, 63*(8), 85–86.
- Rao, S., Collins, H.L., & DiCarlo, S.E. (2002). Collaborative testing enhances student learning. *Advances in Physiology Education, 26*(1), 37-41.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing, 18*(4), 429-462.
- Rinaldi, C. (2006). *In Dialogue with Reggio Emilia: Listening, Researching and Learning*. New York, USA: Routledge.
- Roeber, E.D. (2003). *Assessment models for No Child Left Behind*. Denver, CO: Education Commission of the States.
- Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(1), 119-144.
- Scheerens, J. (2013). The use of theory in school effectiveness research revisited. *School effectiveness and school improvement, 24*(1), 1-38.
- Scheerens, J. (2016). *Educational Effectiveness and Ineffectiveness: A Critical Review of the Knowledge Base*. Dordrecht, The Netherlands: Springer.
- Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- Schmoker, M. (2006). *Results now*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Shepard, L.A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14.
- Shepard, L.A. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 1066-1101). Washington, DC: American Educational Research Association.

- Shepard, L.A. (2007). Formative assessment: Caveat emptor. In C.A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 279–303). Mahwah, NJ: Erlbaum.
- Slater, R.O., & Teddlie, C. (1992). Toward a theory of school effectiveness and leadership. *School Effectiveness and School Improvement*, 3(4), 247-257.
- Stiggins, R.J. (1992). High quality classroom assessment: What does it really mean? *Educational Measurement: Issues and Practice*, 11(2), 35–39.
- Stiggins, R.J. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan*, 86(1), 22–28.
- Stiggins, R., & Chappuis, S. (2005). Putting testing in perspective: It's for learning. *Principal Leadership (High School Ed.)*, 6(2), 16-20.
- Stiggins, R.J., & Chappuis, J. (2008). Enhancing student learning. *District Administration*, 44(1), 42-44.
- Stiggins, R., Arter, J., Chappuis, J., & Chappuis, S. (2006). *Classroom assessment for student learning: Doing it right—using it well*. Merrill Education/ETS college textbook series. Upper Saddle River, NJ: Pearson Education, Inc.
- 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.
- Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. London, UK: Falmer Press.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: teaching, learning and assessment in the classroom*. Buckingham, UK: Open University Press.
- Visscher, A. J., & Coe, R. (2002). *School improvement through performance feedback* Rotterdam: Swets & Zeitlinger.
- William, D., & Thompson, M. (2007). Integrating assessment with instruction: what will it take to make it work? In C.A. Dwyer (Ed.), *The future of assessment: shaping teaching and learning* (pp. 53-82). Mahwah, NJ: Lawrence Erlbaum Associates.
- 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.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45(4), 477-501.