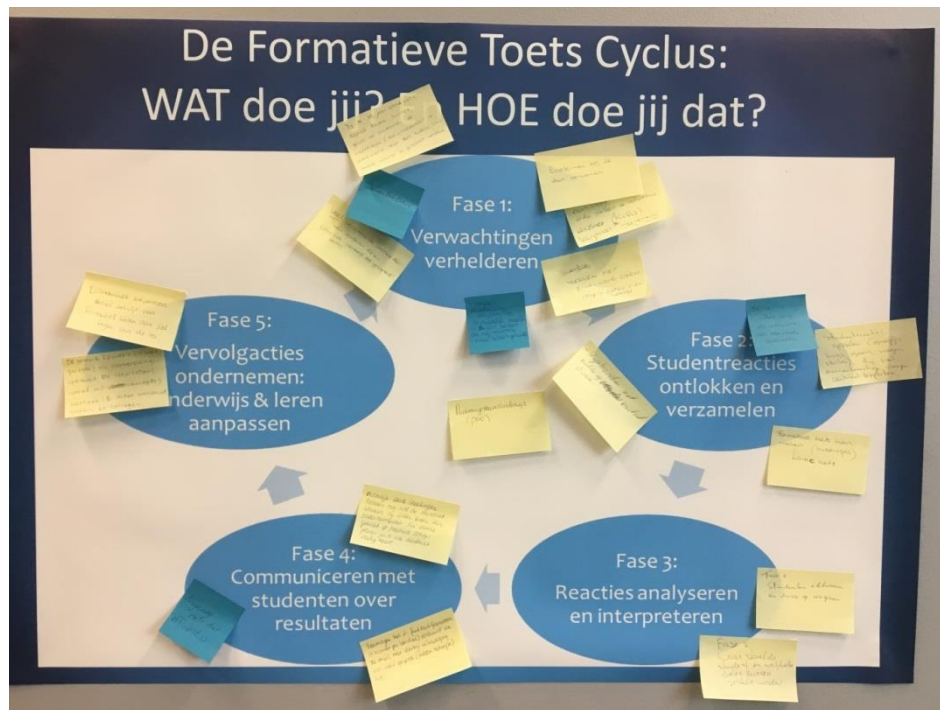


Doelgericht professionaliseren: formatieve toetspraktijken met effect! Wat DOET de docent in de klas?



Overzichtsstudie uitgevoerd met subsidie van het NRO-PPO, dossiernummer 405-15-722

Judith Gulikers, Educatie en competentiestudies, Wageningen Universiteit

Liesbeth Baartman, Lectoraat beroepsonderwijs, Hogeschool Utrecht

Inhoudsopgave

Voorwoord	4
Samenvatting	5
Hoofdstuk 1: Aanleiding en onderzoeksvragen.....	7
Hoofdstuk 2: Theoretische achtergrond	9
Formatief toetsen: Assessment for Learning en Data-based decision making	9
Definiëring op de zeven dimensies van formatief toetsen.....	10
De formatieve toetscyclus: theoretisch en analysekader voor wat de docent DOET in de klas.....	11
Hoofdstuk 3: Onderzoeksmethode	14
Hoofdstuk 4. Resultaten	18
Algemene gegevens.....	18
Onderzoeksvraag 1.....	19
De kwantitatieve resultaten.....	19
De kwalitatieve resultaten.....	20
Typerende docentgedragingen in fase 1: Verwachtingen verhelderen	20
Typerende docentgedragingen bij fase 2: Ontlokken en verzamelen van studentreacties.	23
Typerende docentgedragingen in fase 3: Analyseren en interpreteren van studentreacties	27
Typerende docentgedragingen in fase 4: Communiceren met studenten over resultaten.....	30
Typerende docentgedragingen in fase 5: Vervolgacties ondernemen	31
De gehele Formatieve Toetscyclus en de kleine 2-3-5 cyclus	36
Onderzoeksvraag 2: Studenteffecten en relaties met docentgedragingen	38
Soorten effecten.....	38
Relatie tussen docentgedrag en effecten op studenten	40
Onderzoeksvraag 3: Docentcompetenties voor formatieve toetsgedragingen	41
Doorleefd begrip.....	42
Vakinhoudelijke kennis en learning progressions.	42
Vakdidactisch handelingsrepertoire.	43
Student-docent relatie met student-sturing.	43
Inzetten van technologie voor ondersteuning van het formatieve toetsproces.	44
Onderzoeksvraag 4: Professionaliseringstrajecten en effecten	45
Hoofdstuk 5. Kennisdeeldag	48
Belangrijkste reacties en discussiepunten uit het veld	50
Hoofdstuk 6: Conclusies en discussie	52
Hoofdstuk 7: Limitaties en implicaties	55
Limitaties.....	55
Theoretische implicaties	56
Praktische implicaties: Suggesties voor professionalisering	57
Referenties	59
Referenties meegenomen in de review (N= 106).....	59
Overige referenties.....	63
Appendices	65

Appendix 1. Codeerschema voor codering in Atlas-Ti	65
Appendix 2: overzicht analyse van code "teacher behaviour → effect on student learning" (onderzoeksvraag 2).....	68
Appendix 3. overzicht van professionaliseringstrajecten, FT-fasen en effecten op docenten (onderzoeksvraag 4).....	80
Appendix 4. Posters van de vijf FT-fasen en de professionaliseringsposter	108

Voorwoord

Voor u ligt de rapportage van onze NRO-PPO overzichtsstudie naar “doelgericht professionaliseren. Formatief toetsen met effect: wat doet de docent in de klas?”

Met heel veel plezier hebben wij aan deze studie gewerkt. De focus op de concrete activiteiten van de docent in de klas heeft voor ons een rijkheid aan nieuwe artikelen en inzichten opgeleverd. Concrete Inzichten over wat een docent doet in de klas om effectieve formatieve toetspraktijken te bereiken. Hiervoor hebben we “de formatieve toetscyclus met vijf fasen” gebruikt als analysekader om verschillen in meer en minder effectieve docentgedragingen te duiden en te differentiëren. Vanuit deze concrete gedragingen hebben wij gekeken naar effecten op studenten: welke gedragingen leiden nu tot welke effecten op studenten? Ondanks dat er nog maar weinig empirisch bewijs is voor effecten van formatief toetsen op studenten, laat deze overzichtsstudie toch een heel inspirerend resultaat hierbij zien.

Vanuit de concrete docentgedragingen hebben we vervolgens gekeken naar: wat moet de docent kennen en kunnen om deze gedragingen te kunnen laten zien? Hieruit is een overzicht van cruciale competenties ontstaan die direct gekoppeld zijn aan de uitvoering van verschillende formatieve toetsgedragingen. Als laatste is gekeken naar welke formatieve toetsfasen en onderliggende competenties aan bod komen in professionaliseringstrajecten en welke effecten hiermee worden bereikt op docenten: gaan docenten daadwerkelijk de gewenste formatieve toetsgedragingen meer laten zien? En waar is dan nog winst te behalen?

Na afloop zijn alle resultaten in een interactieve kennisdeeldag bediscussieerd met een gevarieerde groep docenten: wat zien zij in hun eigen formatieve toetspraktijk als ze met de gedragingen en competenties van deze studie daar eens kritisch naar kijken? Welke inspirerende voorbeelden, passend binnen welke fase(n) van de formatieve toetscyclus, hebben zij te delen? En welke onderdelen prioriteren we samen voor verdere professionalisering?

Wij hopen dat de resultaten van deze studie voor u net zo inspirerend zijn dan dat het werken eraan voor ons is geweest. Wij hopen met deze resultaten docenten te inspireren om in hun eigen klas te gaan experimenteren met formatieve toetsgedragingen en zo een stapje dichterbij implementatie van betere formatieve toetspraktijken te komen. Op basis van de resultaten benoemen wij concrete suggesties voor verdere professionalisering en benoemen wij ideeën waar verdere theoretische onderbouwing en onderzoek naar formatief toetsen de vruchten van zou kunnen plukken.

Een paar opmerkingen voor de leesbaarheid: In deze rapportage spreken we van “studenten”, waarmee alle lerenden op alle onderwijsniveaus worden bedoeld. Ook spreken we van “formatief toetsen” als bundeling voor namen als formatieve assessment, evalueren om te leren, of toetsen om van te leren.

Wij wensen jullie heel veel leesplezier!

Judith Gulikers, Educatie en competentiestudies, Wageningen Universiteit

Liesbeth Baartman, Lectoraat beroepsonderwijs, Hogeschool Utrecht

Samenvatting

De belangrijkste aanleiding voor deze reviewstudie is de wereldwijd erkende kernrol die de docent speelt in formatief toetsen, maar het voortdurende probleem om formatief toetsen daadwerkelijk te implementeren in de onderwijspraktijk. Deze reviewstudie onderzoekt welke gedragingen de docent typeren die een formatieve toetspraktijk in de klas weet te realiseren. De focus op de concrete praktijken van de docent in de klas schijnt een nieuw licht op formatief toetsen, namelijk wat de docent in zijn/haar eigen klas kan doen in termen van kleine en grotere gedragingen om een meer formatieve toetspraktijk vorm te geven. De formatieve toetscyclus, bestaande uit vijf fasen gebaseerd op een combinatie van verschillende formatieve toetsmodellen, is gebruikt als theoretisch en analysekader. Deze vijf fasen zijn: 1) verhelderen van verwachtingen, 2) ontlokken van studentreacties, 3) analyseren en interpreteren van studentreacties, 4) communiceren over resultaten met studenten, en 5) vervolgactiviteiten ondernemen. Deze vijf fasen zijn handzaam gebleken om zo concreet mogelijk docentgedrag te duiden en tussen meer of minder effectieve gedragingen te differentiëren. Vier onderzoeksvragen worden hiermee beantwoord:

- 1) Hoe zien effectieve formatieve toetspraktijken in de klas eruit? Met andere woorden, wat DOET de docent die effectieve formatieve toetspraktijken in de klas bereikt?
- 2) Welke effecten op studenten hebben deze toetspraktijken in termen van leeractiviteiten, leerprocessen en/of leeruitkomsten?
- 3) Welke formatieve toetscompetenties heeft een docent nodig om deze effectieve formatieve toetspraktijken in de klas te kunnen vormgeven?
- 4) Aan welke formatieve toetscompetenties (onderzoeksvraag 3) en gedragingen (onderzoeksvraag 1) besteden professionaliseringstrajecten aandacht? Op welke gebieden zijn deze succesvol en waarop (nog) niet?

De literatuursearch bestaat uit twee ronden: ten eerste is een secundaire analyse op de twee voorgaande NRO/NWO reviews naar formatief toetsen (Schildkamp et al., 2014; Sluijsmans et al., 2013) met de inclusiecriteria geformuleerd voor deze studie, waarbij concrete informatie over wat de docent doet in de klas het meest cruciaal was. Dit resulteerde in een selectie van 34 studies. Een aanvullende search in Web of Science, ERIC, Scopus en PsychInfo op voor deze studie geformuleerde zoekcriteria resulteerde na verschillende ronden in 72 aanvullende studies. De 106 studies zijn gecodeerd in Atlas-Ti, kwantitatief en uitgebreid kwalitatief geanalyseerd.

Voor onderzoeksvraag 1 zijn per fase typerende gedragingen gevonden die het onderscheid maken tussen een docent die een meer en minder effectieve formatieve toetspraktijk weet te realiseren. Opvallend hierbij is dat geen enkele studie aandacht besteedt aan de vijf fasen van de formatieve toetscyclus. Een “kleine cyclus” zien we echter wel veelvuldig terug, namelijk de 2-3-5 cyclus: ontlokken van studentreactie (fase 2), analyseren en interpreteren (fase 3), vervolgacties ondernemen (fase 5). Een selectie van opvallende bevindingen per fase was voor fase 1 het werken met ‘learning progressions’, veel informeel “on-the-fly” toetsen in fase 2 door het stellen van goede vragen gericht op begrip en doorvragen in klasdiscussies (2-3-2-3 cyclus), weinig concrete informatie over fase 4 (feedback) en fase 5 die ofwel weinig concreet onderzocht is of voor docenten moeilijk bleek.

Wat betreft de effecten op studenten en de relaties met docentgedragingen, zijn uit de analyses vijf typen effecten te destilleren, namelijk effecten op 1) kennisontwikkeling; 2) schrijven; 3) motivatie/attitude; 4) leerprocessen zoals zelf-regulatie, en 5) studentpercepties of sfeer in de klas. Analyse van de relaties tussen concrete docentgedragingen en studenteffecten laat een trend zien van daadwerkelijk ander docentgedrag in de formatieve toetscyclus wanneer de docenten kennisontwikkeling danwel zelf-regulatie wil stimuleren met formatief toetsen

Uit de concrete docentpraktijken zijn voor onderzoeksvraag 3 noodzakelijke kennis en kunde voor de docent gedestilleerd. De gevonden docentcompetenties zijn dus altijd direct gekoppeld aan gedragingen uit de formatieve toetsfasen en maken daarmee de competenties direct concreet. De gevonden competenties (in de resultaten sectie uitgebreid gekoppeld aan de fasegedragingen) zijn: 1) doorleefd begrip van formatief toetsen hebben; 2) vakinhoudelijke kennis en inzicht in learning progressions kunnen toepassen; 3) een vakdidactisch handelingsrepertoire kunnen inzetten; 4) student-docent relaties expliciet aan de orde kunnen stellen met een focus op student-sturing, 5) technologie kunnen gebruiken ter ondersteuning van het formatieve toetsproces.

Voor de laatste onderzoeksvraag zijn de 43 van de 106 studies bekeken die een professionaliseringsinterventie in zich hebben. In kaart is gebracht welke formatieve toetsfase(n) en -competenties deze trajecten expliciet aanspreken en welke effecten op docenten dit opleverde. Dit overzicht biedt een rijkheid aan informatie over in welke formatieve toetsfasen docenten goed geprofessionaliseerd worden en deze professionalisering ook werkt en waar niet. Enkele saillant bevindingen hieromtrent zijn dat geen enkel traject aandacht besteedt aan de gehele formatieve toetscyclus, professionalisering in doelen en learning progressions (fase 1) voorwaardelijk voor de gehele toetscyclus lijkt te zijn, trajecten die twee of meer fasen in samenhang professionalisering (alignment) meer succesvol zijn, en professionalisering in fase 5 nog weinig succesvol. Docenten zijn veelal nog handelingsverlegen om goede didactische vervolgacties te kiezen en vorm te geven.

Hoofdstuk 5 beschrijft de activiteiten en belangrijkste reacties van een landelijke kennisdeeldag met het veld. Op deze dag zijn 52 deelnemers, bestaande uit docenten van alle onderwijsniveau, onderwijsadviseurs, beleidsmedewerkers en vertegenwoordiging van Ministerie van OCW, aan de slag gegaan met de resultaten. Via verschillende werkvormen hebben zij kennisgenomen van de resultaten (zie ook poster in appendix 4), kritische gekeken naar hun eigen formatieve toetspraktijken, elkaar geïnspireerd met concrete voorbeelden en bediscussieerd waarop verdere professionalisering het meest noodzakelijk is. Door middel van een kaartspel hebben deelnemers hun eigen formatieve toetspraktijken geëxpliciteerd en met elkaar gedeeld. Zij hebben een plaatje van zichzelf gemaakt op de vijf fasen van de formatieve toetscyclus (Wat doe ik al goed, wat wil ik leren en wat is mijn concrete en inspirerende voorbeeld?). De bevindingen van deze dag zijn meegenomen om vooral de praktische implicaties van deze studie te formuleren en de bundeling van goede voorbeelden is een deelbaar praktisch resultaat van deze reviewstudie (*nb. Op te leveren in maart 2017*).

Hoofdstuk 6 bediscussieert de belangrijkste conclusies van deze studie. Hoofdstuk 7 benoemt enkele kanttekeningen waarmee bij de conclusies rekening gehouden moet worden en formuleert theoretische en praktische implicaties. Een interessante theoretische implicatie is dat deze studie zich bevindt op het snijvlak van toetsonderzoek en instructieonderzoek. De formatieve toetsfasen kwamen ook geregeld terug in artikelen die vanuit een meer instructie-georiënteerde stromingen geschreven waren, bijvoorbeeld scaffolding of noticing. Theorie en onderzoek naar formatief toetsen kan zeker verrijkt worden door aanknopingspunten met deze andere onderzoeksdomeinen aan te gaan. De praktische implicaties zijn gericht op welke nieuwe inzichten deze studie oplevert voor professionalisering van docenten. Hierbij verdient, naast andere punten, met name fase 5 expliciete aandacht waarbij expliciete aandacht uit moet gaan naar het werken aan een vakdidactisch handelingsrepertoire om beter vorm te geven aan fase 5 (vervolgacties ondernemen).

Hoofdstuk 1: Aanleiding en onderzoeksvragen

Weer een review naar formatief toetsen? Zowel internationaal als nationaal zijn er al vele reviewstudies naar formatief toetsen uitgevoerd. Sommige gaan al enkele tientallen jaren terug, andere zijn zeer recent. Sommige zijn vooral gebaseerd op theoretische verhandelingen over formatief toetsen, andere zijn meta-studies waarin empirische resultaten worden blootgelegd en vergeleken. Ook zijn er verschillende reviews die vooral bestaan uit een kritische review van de bevindingen van eerdere reviews. Kortom: in de literatuur is veel gaande als het gaat om formatief toetsen. Echter, nog steeds is de theoretische definiëring en onderbouwing van wat formatief toetsen is onduidelijk, komt implementatie in de klas nog niet goed van de grond, is professionalisering op dit thema velerlei en is het bewijs voor effecten van formatief toetsen op studenten mager.

Dit zijn ook de conclusies van twee recent uitgevoerde NRO-PPO review studies naar formatief toetsen (Schildkamp et al, 2014; Sluijsmans et al., 2013). De review van Sluijsmans en collega's legt de variatie in definiëring van formatief toetsen bloot en identificeert acht effectieve methodieken voor formatief toetsen. Deze review benoemt tevens drie nog open liggende onderzoeksterreinen: versterking van de theorie van formatief toetsen, meer effectstudies naar formatief toetsen, en condities voor duurzame implementatie. De review van Schildkamp en collega's bouwt met name voort op dit laatste onderzoeksterrein. Deze review identificeert de (rand)voorwaarden voor toets, toetsinstrument, toetsproces, rol van docent, rol van student en de context noodzakelijk voor formatieve toetsing om effect op studenten te hebben. Ook deze studie concludeert dat er nog weinig bewijs is voor daadwerkelijk effect van formatief toetsen op studenten en roept op tot effectstudies. Beide reviews concluderen bovendien dat de docent en de studenten in de klas de cruciale schakels zijn in het formatieve toetsproces: zo gaat het erom hoe zij een methodiek inzetten en niet om de methodiek zelf. Beide reviews benadrukken sterk de behoefte aan docentenprofessionalisering. Hierbij gaat het om vragen als: "Wat moeten docenten kennen en kunnen om formatief toetsen vorm te geven?" en "Wat zijn effectieve vormen van professionalisering op dit gebied?".

Ook de onderwijspraktijk laat zien dat het implementeren van formatief toetsen door docenten in de klas geen sinecure is. Om allerlei redenen blijkt dit erg lastig (zie bijvoorbeeld alle randvoorwaarden benoemd door Schildkamp et al., 2014). Mede hierdoor is het effect op studenten moeilijk meetbaar en moeilijk te bestuderen.

Nationaal en internationaal worden kleine en grootschalige interventies uitgevoerd gericht op het professionaliseren van docenten, teams en schoolleiders in bepaalde aspecten van formatief toetsen. Deze professionaliseringsactiviteiten kunnen gericht zijn op zaken als: "hoe geef ik goede feedback?", "hoe ontwerp ik een beoordelingsrubric?", "hoe betrek ik studenten bij self/peer-assessment", "hoe analyseren wij data van onze studenten?", "hoe creëren we een formatieve toetscultuur in onze school" en zo meer. Het implementeren van formatief toetsen in een school vraagt aanpassingen op schoolniveau, denk aan onderwijs/toetsvisie en cultuur, maar de docent in de klas blijft de cruciale schakel in het potentieel bewerkstellingen van effect bij studenten. De laatste jaren komt er steeds meer bewijs voor een actieonderzoek gerichte vorm van professionalisering waarin deelnemers vanuit en in hun eigen praktijk werken aan het verbeteren van deze praktijk. Het onderzoeken en daarmee verbeteren van de eigen praktijk lijkt het meest vruchtbaar voor het daadwerkelijk implementeren van (aspecten van) formatief toetsen.

De vraag is dan nu hoe we de theorie én de praktijk van formatief toetsen daadwerkelijk een stap verder kunnen helpen. Recente (review)studies roepen steeds meer op tot het bekijken, bestuderen en beschrijven van concrete praktijken in de klas: wat DOET een docent die aan formatief toetsen doet, of zegt te doen? (Antoniou & James, 2014; Sluijsmans, Joosten-ten Brinke, & Van der Vleuten, 2013). Deze studies roepen expliciet op om meer aandacht besteden aan een kritische analyse van deze concrete praktijken van docenten en actieonderzoek hierin samen met de docenten. Vanuit deze inzichten kunnen we echt verder bouwen aan theorie over en bewijs voor formatief toetsen (Antoniou & James, 2014; Torrance, 2001).

Dit is de aanleiding voor deze voorliggende overzichtsstudie. In deze studie zijn concrete formatieve toetspraktijken van de docent in de klas (wat DOET de docent in de klas) het startpunt van analyse. Vanuit deze concrete formatieve toetspraktijk wordt gekeken naar: wat doet de docent? Wat doet hierdoor de student? Wat is vervolgens het leereffect op de student? Hiermee krijgen we zicht op wat concrete effectieve formatieve toetspraktijken in de klas inhouden. Van hieruit wordt gekeken naar de vraag wat docenten moeten kennen en kunnen om deze activiteiten in de klas vorm te geven. En als we dat weten, kunnen we ook concrete, gerichte aanbevelingen doen over manieren waarop docentenprofessionalisering plaatsvinden en onderzoeken in hoeverre bestaande professionaliseringstrajecten hier al dan niet aandacht aan besteden en waar dus verbeteringen te behalen zijn.

De concrete onderzoeksvragen van deze overzichtsstudie zijn hiermee als volgt samen te vatten:

- 5) Hoe zien effectieve formatieve toetspraktijken in de klas eruit? Met andere woorden, wat DOET de docent die effectieve formatieve toetspraktijken in de klas bereikt?
- 6) Welke effecten op studenten hebben deze toetspraktijken in termen van leeractiviteiten, leerprocessen en/of leeruitkomsten?
- 7) Welke formatieve toetscompetenties heeft een docent nodig om deze effectieve formatieve toetspraktijken in de klas te kunnen vormgeven?
- 8) Aan welke formatieve toetscompetenties (onderzoeksvraag 3) en gedragingen (onderzoeksvraag 1) besteden professionaliseringstrajecten aandacht? Op welke gebieden zijn deze succesvol en waarop (nog) niet?

Dit rapport legt verslag van een systematische literatuurreview naar deze vragen, waarbij de concrete formatieve toetspraktijken van de docent in de klas de basis vormden voor de analyse. Een theoretisch onderbouwde formatieve toetscyclus van 5 fasen (zie hoofdstuk 2) is de onderlegger voor de analyse. Door middel van Atlas-ti zijn alle gevonden bronnen kwantitatief en kwalitatief geanalyseerd (hoofdstuk 3) om antwoorden te formuleren op de vier onderzoeksvragen (hoofdstuk 4).

Vervolgens zijn deze resultaten per fase van de formatieve toetscyclus vertaald naar praktische output voor docenten in de vorm van posters en een kaartspel. Hierin is te vinden: 1) wat doet een docent die effectieve formatieve toetspraktijken laat zien per fase; 2) wat blijken docenten nog veel te weinig te laten zien; 3) wat moet een docent kennen en kunnen om deze toetspraktijken vorm te geven in de klas, en 4) welke vragen stel je jezelf en elkaar om kritisch te reflecteren op de eigen formatieve toetspraktijk. Deze praktische output is gebruikt in een kennisdeeldag met een groot aantal docenten werkzaam op verschillende onderwijsniveaus. Hiermee zijn de bevindingen uit de overzichtsstudie verfijnd en gevalideerd voor de praktijk. Hoofdstuk 5 beschrijft de belangrijkste reacties en discussiepunten van deze kennisdeeldag en voedt hiermee de praktische implicaties, waaronder toekomstige professionalisering, van deze studie.

Hoofdstuk 6 trekt conclusies over de onderzoeksvragen op basis van de theoretische en praktische dataverzameling. En hoofdstuk 7 bediscussieert de studie. In dit hoofdstuk worden eerst een paar kanttekeningen bij de studie geplaatst die meegenomen moeten worden bij het interpreteren van de gegevens. Daarna formuleren wij hier de theoretische bijdrage van deze studies en de praktische implicaties voor de onderwijspraktijk. Hierin wordt expliciet stil gestaan bij wat op basis van deze onderzoeksresultaten gezegd kan worden over toekomstige professionalisering van docenten.

Hoofdstuk 2: Theoretische achtergrond

Formatief toetsen, oftewel toetsen om *van de* leren, krijgt al jaren veel aandacht. Vooral omdat ervan verwacht wordt dat het een positief effect heeft op het leren van studenten. Dit hoofdstuk gaat kort in op hoe en van waaruit de ideeën over formatief toetsen ontstaan zijn en waarom het moeilijk is om in de praktijk vorm te geven en te implementeren. Van daaruit beschrijven we hoe wij in deze voorliggende studie naar formatief toetsen hebben gekeken. Sluijsmans, Joosten-Ten Brinke en van der Vleuten (2013) rapporteren de brede variatie aan definities voor formatief toetsen. Geen enkele definitie is fout, maar de definitie die je kiest heeft wel (grote) invloed op hoe een studie met het concept omgaat. Bij een reviewstudie als deze, is de gekozen definiëring van grote invloed op de zoek- en inclusiecriteria en daarmee dus op de uiteindelijke geselecteerde artikelen en de daaruit voortkomende resultaten.

Formatief toetsen: Assessment for Learning en Data-based decision making

Twee grote stromingen hebben een belangrijke rol gespeeld bij de opkomst van formatief toetsen. De “assessment for learning” (AfL) stroming en de “data-based decision making” (DBDM) stroming (Van der Kleij, Vermeulen, Schildkamp, & Eggen, 2015). De AfL stroming is vooral ontstaan en uitgewerkt in Brittannië, door de Assessment Reform Group en namen als bijvoorbeeld Black, Wiliam, Stobart, James en Gipp. De DBDM stroming komt vooral uit de Verenigde Staten en krijgt in Nederland inmiddels meer bekendheid en populariteit als “Opbrengst Gericht Werken” (Inspectie van het Onderwijs, 2011). Beide stromingen hebben belangrijke verschillen, maar hebben gemeen dat ze bedoeld zijn om student data (dwz. toetsgegevens, prestaties, antwoorden, oefenopgaven ed.) doelgericht te gebruiken om het leren van de student te verbeteren. AfL is vooral gericht op het leerproces in de klas, de interactie tussen student en docent en doelt vaak op het stimuleren van leerprocessen als zelf-sturing, zelf-regulatie, of leermotivatie (bv. Clark, 2012; De Kleijn, Prins, Lutz, van Look, & van Tartwijk, 2015; Sluijsmans et al., 2013). DBDM speelt daarentegen meer op school/vak niveau en is vooral gericht op het verhogen van leerresultaten, vaak gemeten via landelijke, gestandaardiseerde metingen, zoals de leerlingvolgsystemen van Cito. De doorlooptijd van DBDM is dan ook vaak groter en langer dan bij AfL en gebeurt voor groot deel buiten de klas (van der Kleij et al., 2015).

Assessment for Learning benadrukt de cruciale rollen van de docent en studenten in de klas in het proces van formatief toetsen en verwacht grootse effecten op studenten, maar veel studies zijn vooral theoretisch en een grote set aan reviewstudies laat zien dat de bewezen effecten mager zijn, elkaar tegenspreken of bestrijden, of gebaseerd op zwakke onderzoeksdesigns (bv. Bennett, 2011; De Kleijn et al, 2015; Kingston & Nash, 2011). Ook in de empirische studies zijn we een grote variatie aan studies, met name doordat deze studies formatief toetsen op geheel andere wijze definiëren: De wat meer grootschalige onderzoeken met een voor- en nameting (en soms een controlegroep) zijn vaak geplande en georganiseerde vormen van formatief toetsen (bv. “embedded assessment” Yin et al., 2008; Ayala et al., 2008) en gericht op de gemakkelijker meetbare leeruitkomsten zoals scores op landelijke (state-wide) tests (b.v.). Studies gericht op de moeilijker meetbare leerprocessen en –uitkomsten, waarbij vooral veel aandacht is voor zelf-regulerende vaardigheden als plannen, monitoren, reflecteren en daarmee zelf eigen leren sturen (Clark, 2012) of motivatie (de Kleyn, Prins et al), zijn schaars, kleinschalig en beschrijvend. Dit geldt ook voor studies die expliciet zijn gericht op het betrekken van studenten bij het formatieve toetsproces en het bevorderen van eigenaarschap (goede voorbeelden zijn Kearney, 2013; Lorente & Kirk, 2013). Door deze grote variatie in ‘soorten’ studies en definities van formatief toetsen die eronder liggen, is het moeilijk om studies te vergelijken en een uitspraak te doen over welke docentactiviteiten daadwerkelijk bijdragen aan het leren van studenten (Bennett, 2011).

DBDM daarentegen wordt ondersteund door meer effectstudies met sterkere longitudinale onderzoeksdesigns (e.g. Schildkamp, Lai & Earl, 2014) maar deze zijn vooral gericht op gestandaardiseerde leerresultaten en leerlingvolgsystemen (bv. Rekeken, taal) en zijn minder gericht op het leerproces tussen docent en student. DBDM processen vinden voor een groot deel buiten de klas plaats, en zeggen daarmee niet altijd veel over wat de docent doet in de klas. Echter, recente onderzoeken in de context van de sterke opkomst van Opbrengst

gericht werken in Nederland laten zien dat ook in OGW de docent juist een cruciale sleutelrol speelt (Vermaas, 2011)

Voor deze Overzichtsstudie hebben we gebruik gemaakt van beide stromingen omdat ze beide (1) ze beide gericht zijn op het gebruiken van studentgegevens om het leren te optimaliseren; (2) benadrukken dat de docent een cruciale rol heeft; (3) laten zien dat implementatie in de klas moeilijk is en blijft; (4) ze werken vanuit een cyclische gedachte (zie hieronder), en ze (5) elkaar kunnen aanvullen in het geven van goed en gedegen inzichten in wat een docent DOET in de klas om daadwerkelijke effecten te bereiken op leerprocessen en leeruitkomsten van studenten.

Definiëring op de zeven dimensies van formatief toetsen

Sluijsmans en collega's (2013) benoemen 7 dimensies waarop studies en praktijken van formatief toetsen verschillen. Omdat deze overzichtsstudie aansluit op de voorgaande PPO-NRO reviewstudies (Sluijsmans et al., 2013; Schildkamp et al., 2014) wordt de definiering van formatief zoals gehanteerd in deze overzichtsstudie vergeleken met de 7 dimensies die Sluijsmans et al (2013) definiëren. In onderstaande tabel staat beschreven welke variatie aan formatieve toetsing in deze overzichtsstudie zijn meegenomen.

De dimensie "classroom assessment versus formatieve feedback" gaat over wat allemaal onder formatief toetsen valt. De ene kant van de dimensie (Classroom assessment) zie alle toetsing in de klas als formatief toetsen (dus ook de toetsen die wel een cijfer of beoordeling krijgen), terwijl de andere kant van de dimensie alleen iets formatief noemt als het leidt tot feedback gericht op het stimuleren van zelf-regulerende vaardigheden

De dimensie "bevorderen van leren van studenten versus het leren van docenten" gaat over de opvatting of studentgegevens gebruikt moeten worden door studenten om hun leren verder te sturen aan de ene kant van de dimensie tot of dat zij gebruikt moeten worden door docenten om op hun lespraktijk te reflecteren en deze aan te passen.

De dimensie "instrument versus proces" varieert van formatief toetsen zien als het inzetten van een toetsinstrument, methode of strategie tot formatief toetsen zien als een proces tussen student en docent onafhankelijk van welk instrument hierbij gebruikt wordt.

De dimensie "summatief versus formatief" gaat over op welke manier en hoe intensief een formatief toets gebruikt wordt om het leren te stimuleren. Dit varieert van een formatieve toets zonder leerwaarde en feedback, waarbij docenten een toets formatief noemen maar geen aandacht besteden aan het gebruiken van de resultaten om feedback te geven of het leren te stimuleren. De andere kant van de dimensie is een formatieve toets gebruiken om zelf-regulerende vaardigheden bij studenten verder te ontwikkelen.

De dimensie "georganiseerde inbedding versus on-the-fly" maakt het verschil tussen geplande, georganiseerde en vooraf (door de docent/school) bedachte formatieve toetsing aan de ene kant tot volledig spontane, geïmproviseerde formatieve toetsen in de klas op het moment dat zich bepaalde studentreacties voordoen. Deze dimensie maakt eigenlijk het onderscheid tussen formeel en informeel formatief toetsen.

De dimensie "docent versus student-gestuurd" gaat over wie (de docent, de student, of gezamenlijk) de inhoud, invulling en gebruik van de toetsing en resultaten stuurt en bepaalt.

De laatste dimensie "stap-voor-stap versus volledige inbedding in de leercyclus" gaat over de mate waarin formatieve toetsing losstaand is, of een meer geïntegreerd en automatisch onderdeel van het onderwijsleerproces.

In onderstaande tabel beschrijven we per dimensie hoe deze studie formatief toetsen definieert en welke studies daarmee wel of juist niet in de scope van deze overzichtsstudie vallen.

Dimensies van Sluijsmans et al (2013)	Deze studie definieert formatief toetsen als.....
Classroom assessment vs formatieve feedback	Alle vormen van formele en informele toetsing in de klas die gebruikt worden om zicht te krijgen op het leren van studenten om op basis hiervan het leren van de studenten te bevorderen en/of onderwijsactiviteiten aan te passen.
Bevorderen van leren van student of leren van leraren	Focus ligt op de docent, maar het doel van de studies in deze review is een effect teweeg te brengen bij de student: op welke manier ontlokt, analyseert en gebruikt de docent data van studenten om hiermee het leren van de student en zijn/haar eigen onderwijs aan te passen.
Instrument versus proces	Proces. Zicht krijgen op welke activiteiten een docent uitvoert in de klas. Het gebruiken van een instrument kan hier onderdeel van zijn (fase 2 van de formatieve toetscyclus, zie hieronder).
Summatief versus formatief	Het gaat expliciet om toetsing die het leren van studenten blootlegt en stimuleert. Deze studie legt hierbij geen beperking op aan het soort leereffect of leeruitkomst dat wordt gestimuleerd en ook niet op de manier waarop dat effect gemeten wordt. Dat kan dus variëren van effecten op (kennis)toetsresultaten tot effecten op leergedrag of motivatie en zelfregulerende vaardigheden.
Georganiseerde inbedding versus on the fly	Deze studie kijkt naar het gehele spectrum, mits het gaat over activiteiten die de docent doet in de klas.
Leraarsgestuurd versus studentgestuurd	Beide, mits de studies in ieder geval informatie geven over wat de docent doet. Dit gebeurt op eigen initiatief van de docent of aangestuurd door studenten.
Stap voor stap versus volledige inbedding in de leercyclus	Deze studie benadert formatief toetsen als een proces van vijf fasen (zie hieronder). De geselecteerde studies moeten aandacht besteden aan minimaal een van deze fasen.

De formatieve toetscyclus: theoretisch en analysekader voor wat de docent DOET in de klas

Omdat deze overzichtsstudie zich richt op “wat *doet* de docent in de klas die goede formatieve toetspraktijken bereikt” gaat het inherent om een formatief toetsproces. De meest bekende beschrijving van het formatieve toetsproces komt uit de Assessment for Learning traditie en is beschreven door Dylan Wiliam (2011) in drie kernvragen. Hattie en Timperley (2007) gaan uit van dezelfde drie kernvragen en koppelen dit aan activiteiten van feedup, feedback en feedforward. Ook Sluijsmans en collega’s (2013) gebruiken deze drie fasen in hun PPO-NRO review. De drie kernvragen zijn:

1. Waar gaat de student naartoe? Dit noemen Hattie en Timperly ook wel feedup
2. Waar staat de student nu? Dit noemen Hattie en Timperly ook wel feedback
3. Hoe komt de student naar de gewenste situatie / hoe nu verder? Dit noemen Hattie en Timperly ook wel feedforward.

De onderzoeksstroming gericht op Data-based Decision Making (DBDM) beschrijft een vergelijkbare cyclus waarbij de nadruk ligt op het gebruik van data ten behoeve van het optimaleren van de leeruitkomsten van studenten. Lai & Schildkamp (2014) geven aan dat ook in deze traditie verschillende auteurs verschillende stappen benoemen, maar dat de kern van het OGW proces bestaat uit:

- A. Benoemen van een doel
- B. Verzamelen van data
- C. Analyseren van data
- D. Interpretieren van data
- E. Komen tot actie: op basis van interpretaties beslissingen nemen en vervolgstappen zetten.

De stappen A en B komen goed overeen met de stappen 1 en 2 uit de Assessment for Learning traditie (Wiliam, 2011). Echter de OGW traditie is veel specifiek in het van analyseren en interpreteren van data (stap C en D).

Het verschil in stap 3 en E lijkt vooral te zitten in dat bij Wiliam (2011) en Hattie & Timperley (2007) de nadruk op acties door de student ligt, terwijl het vanuit de OGW traditie veelal gaat om aanpassingen die de docent/school doet.

Antoniou en James (2014, zie Figuur 1) beschrijven het formatieve toetsproces aan de hand van vijf activiteiten die de docent doet in de klas. Deze activiteiten sluiten ook aan bij beide bovenstaand faseringen maar zijn explicieter vormgegeven in activiteiten van de docent in de klas als analysekader om docenthandelen te observeren. De activiteiten “communication of expectancies and success criteria” van Antoniou en James (2014) sluit bijvoorbeeld aan bij stap 1 in de AfL traditie: beide zijn gericht op het benoemen en helder maken waar de student naar toe werkt.

Figuur 1: de 5 formatieve toetsactiviteiten van Antoniou & James (2014)

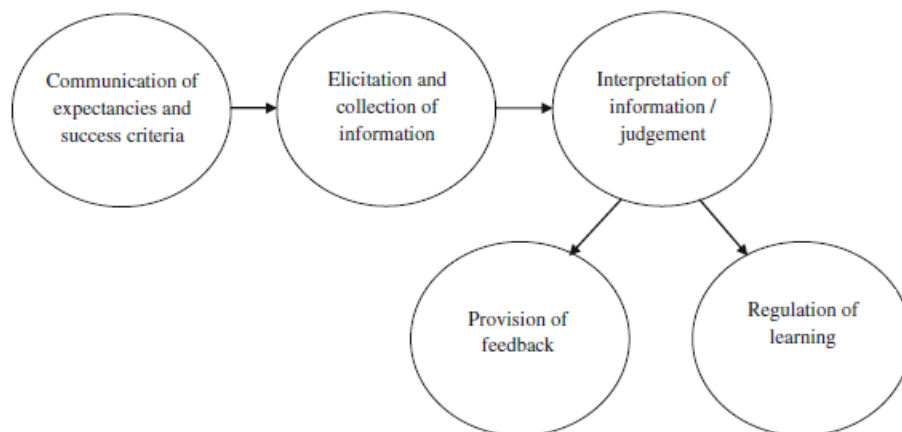


Fig. 1 Theoretical framework of formative assessment

Het activiteitenmodel van Antoniou & James sluit het meest direct aan op het doel van deze overzichtsstudie: in kaart brengen wat een docent in de klas concreet doet (of niet doet). Het activiteitenmodel maakt het tevens mogelijk om te onderzoeken wat de student doet in relatie tot de docentactiviteiten (al dan niet in reactie op wat de docent doet). Dit biedt dus directere handvatten om docenthandelen en studenthandelen aan elkaar te verbinden. Deze vijf activiteiten combineren bovendien zowel bij de drie kernvragen van Assessment for Learning als de vijf stappen van OGW. Wat echter mist in de representatie van Antoniou en James is het cyclische proces van formatief toetsen (e.g. Ruiz-Primo & Furtak, 2007; Sluijsmans et al., 2013). Een mooi voorbeeld van dit cyclische proces voor formatief toetsen in de klas is beschreven door Ruiz-Primo en Furtak in de ESRU-cyclus (zie Figuur 2). Deze cyclus staat voor Elicit – Respond – Recognize – Use. De docent ontlokt een reactie bij de student (elicit), de student reageert (respond), de docent analyseert de student reactie(s) en checkt deze (recognize), en vervolgens gebruikt de docent deze student reactie(s) om een gepaste vervolgactie te ondernemen

Figuur 2: de ESRU-cyclus voor formatief toetsen van Ruiz-Primo & Furtzak (2006).

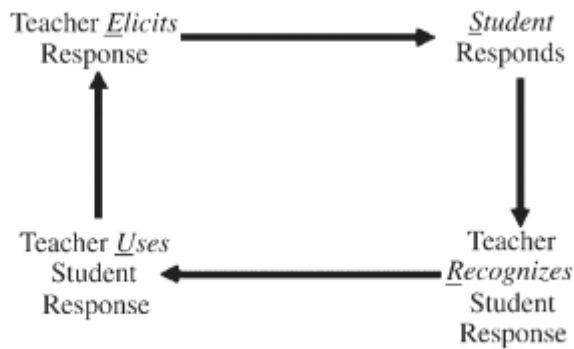
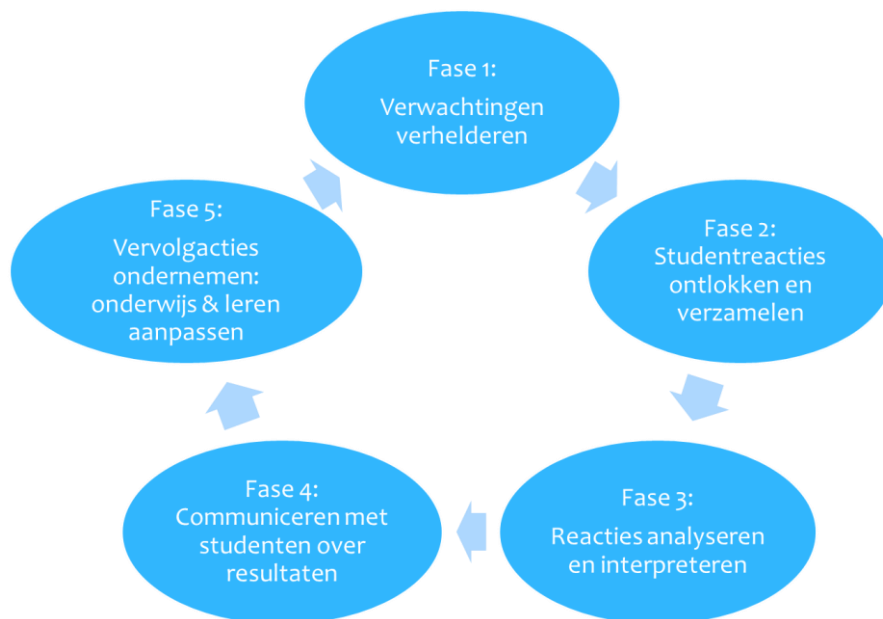


Figure 1. The ESRU model of informal formative assessment.

In deze cyclus van Ruiz-Primo & Furtak (2007) mist echter de belangrijke fase van verwachtingen, leerdoelen en/of succescriteria verhelderen; een fase die in zeer veel formatieve toetsliteratuur onlosmakelijk aan formatief toetsen verbonden is. Om deze redenen vormen de vijf activiteiten van Antoniou en James (2014) gecombineerd met de vier docentactiviteiten en het cyclische idee van de ESRU-cyclus van Ruiz-Primo en Furtak de basis van het theoretische kader en analysekader van deze overzichtsstudie. Dit leidt tot onderstaande formatieve toetscyclus (FT-cyclus), gericht op wat de docent DOET in de klas (Figuur 3). Deze formatieve toetscyclus vormt de basis voor deze overzichtsstudie.

Figuur 3. Theoretisch en analysekader van deze overzichtsstudie: de formatieve toetscyclus.

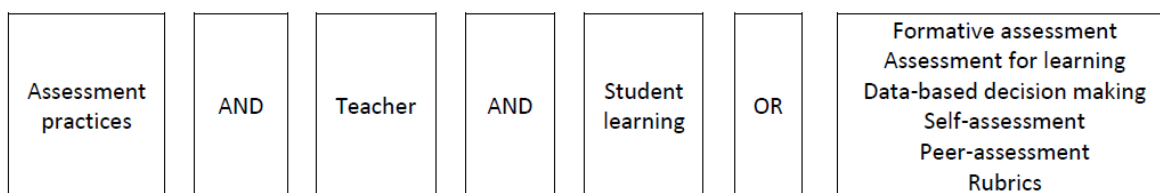


Deze fasen (de FT-fasen) geven voldoende ruimte en toch ook voldoende afbakening om variaties in concreet docenthandelen, studenthandelen en effecten van deze twee op elkaar concreet te maken en in kaart te brengen.

Hoofdstuk 3: Onderzoeksmethode

Deze studie bestaat uit een systematische literatuurreview in twee rondes en is uitgevoerd volgens de stappen van Petticrew en Roberts (2006). De eerste ronde was een secundaire analyse van de twee voorgaande PPO-NRO reviews naar formatief toetsen (Schildkamp et al., 2014; Sluijsmans et al., 2013) op basis van de inclusiecriteria geformuleerd voor deze overzichtsstudie. Een tweede ronde, bestaande uit een eigen aanvullende search in databases ERIC, PsychInfo, Scopus en Web of Science, is gedaan via de zoektermen in Figuur 4 waarbij de concrete assessment practices (wat de docent DOET in de klas) centraal stonden:

Figuur 4: zoektermen voor de literatuursearch



De inclusiecriteria die zijn gehanteerd voor de secundaire analyse van de twee voorgaande PPO-NRO reviews en voor het selecteren van artikelen uit de databases waren:

1. Gepubliceerd in peer-reviewed tijdschrift of boek
2. Focus op formatief toetsen in de schoolse context (ongeacht onderwijsniveau/laag)
3. Empirisch onderzoek
4. Focus op klas-niveau (niet school)
5. Heldere beschrijving van wat doe docent DOET in de klas met betrekking tot een of meerdere fase van de formatieve toetscyclus
6. Resultaten over effecten op studenten in termen van leeruitkomsten en/of leerprocessen

Als eerste zijn de overlappende artikelen uit beide reviews (Schildkamp et al., 2014; Sluijsmans et al., 2013) gehaald. Dit waren er drie. Na een eerste ronde, waarbij de abstracts van de beide reviews zijn bekeken, bleek dat maar een klein deel van de studies daadwerkelijk resultaten op studenten beschreef (inclusiecriterium 6), terwijl in veel van deze studies wel aandacht was voor concrete beschrijvingen van docenthandelen (inclusiecriterium 5). Daarom zijn artikelen met score 5 (stap 8) en 6 (stap 1) beide in zijn totaliteit bekeken op basis van de inclusiecriteria. Als deze artikelen concrete beschrijvingen gaven over wat de docent DOET in de klas (inclusiecriterium 5) zijn ze meegenomen in deze studie, ongeacht of ze wel of neit informatie geven over het effect op studenten (inclusiecriterium 6). Tabel 1 geeft een overzicht van dit proces

Tabel 1: selectieproces van artikelen uit voorgaande PPO-NRO reviews naar formatief toetsen.

Stap	Beslisregel	Sluijsmans et al (exclusief de reviews) Totaal aantal artikelen in review = 32	Schildkamp et al totaal aantal artikelen in review = 74
1	Voldaan aan alle inclusiecriteria	16	25
2	aan geen enkel inclusiecriterium voldaan	1	13
3	niet voldaan aan inclusiecriterium 1	0	0
4	niet voldaan aan inclusiecriterium 2	0	0
5	niet voldaan aan inclusiecriterium 3	11	0
6	niet voldaan aan inclusiecriterium 4	1	12

7	niet voldaan aan inclusiecriteria 5	2	9	
8	niet voldaan aan inclusiecriteria 6	3	9	
9	niet te beoordelen op basis van abstract	7	3	
10	geen abstract)	1	1	
11	Totaal aantal opgenomen na lezen abstracts	16 + 3 + 7	25 + 9 + 3	Stap: 1 + 8 + 9
12	Totaal over na grondig scannen artikelen	14	20	34 artikelen in totaal

De tweede search op basis van de zoekcriteria voor deze specifieke overzichtsstudie resulteerde in eerste instantie in 304 hits. Deze artikelen zijn in verschillende ronden gescreend om te komen tot een uiteindelijke selectie. Dit proces is te zien in tabel 2.

Tabel 2. Selectieprocedure van artikelen uit tweede literatuursearch.

Totale search	Beslisregel	304
delete stap 1	dubbelingen + na scannen titel niet relevant	68
delete stap 2	na lezen abstract niet voldoen aan inclusie criteria	106
delete stap 3	niet te verkrijgen	21
delete stap 4	na scannen artikel niet relevant mbt wat de docent concreet DOET in de KLAS	37
Uiteindelijke aantal opgenomen in review	geselecteerde en in atlas ingevoerde bestanden	72

In totaal bleven er 106 (34 uit de twee eerdere NRO-reviews +72 artikelen uit de aanvullende search) artikelen over die zijn meegenomen in deze overzichtsstudie en zijn ingevoerd en gecodeerd in Atlas-TI. De artikelen zijn *als geheel* gecodeerd op algemene gegevens: onderwijsniveau, vakgebied van het onderwijs, werelddeel, professionaliseringsinterventie gericht op formatief toetsen (ja/nee), student leersuitkomst gemeten (ja/nee), onderzoeksmethodologie, welke van de vijf fasen van de formatieve toetscyclus worden expliciet aangesproken in dit artikel.

Vervolgens werd in de tekst *op fragmentniveau* gecodeerd. Hierin lag de focus op tekstfragmenten die gaan over docenthandelen of studenthandelen op een van de vijf formatieve toetsfasen (gespecificeerd in verschillende concrete docenthandelingen en studenthandelingen, zie bijlage). Hierin werd zowel gecodeerd wat de docent/student WEL doet, maar ook wat hij/zij NIET doet. In dit laatste geval gaat het om activiteiten die in de betreffende interventiestudie wel bedoeld zijn, maar die niet lukken. Bijvoorbeeld, docenten worden getraind om doelen expliciet te maken richting studenten (In dit geval wordt Phase 1 aan het gehele artikel gehangen), maar in de resultaten blijkt dat docenten dit niet doen (in dit geval wordt in de resultaten sectie NOT phase 1, teacher gecodeerd). Deze codes zijn het meest gespecificeerd en zullen antwoord geven op onderzoeksvraag 1.

Vanuit deze gedetailleerde beschrijvingen van wat de docent doet in de klas worden fragmenten ook gecodeerd op effecten van bepaald docenthandelen op studenthandelen of studentuitkomsten (onderzoeksvraag 2, code “teacher effect on student learning”). Daarnaast geven deze gedetailleerde docenthandelingen informatie over wat de docent moet kennen en kunnen om deze handelingen uit te voeren. Tekstfragmenten die hierover gingen werden gecodeerd als “teacher competencies for formative assessment” (onderzoeksvraag 3). Het specifieke codeboek is te vinden in Appendix 1.

Tien artikelen zijn door beide onderzoekers onafhankelijk gecodeerd met als doel het codeboek te verfijnen en gelijkwaardig te interpreteren. Daarna zijn de artikelen verdeeld en door een van beide onderzoekers gecodeerd in Atlas-ti.

Analyse

De artikelen zijn kwantitatief en kwalitatief geanalyseerd om antwoorden te vinden op de vier onderzoeksvragen

Ten eerste zijn de algemene gegevens kwantitatief in kaart gebracht. Dit geeft een beeld waar de studies vooral vandaan komen (welke landen, welke onderwijsniveaus en vakgebieden), welke formatieve toetsfasen vooral bestudeerd worden, welke onderzoeksmethodologie wordt gebruikt, hoe vaak leeruitkomsten worden gemeten (ongeacht type leeruitkomst en zonder strenge eisen zijn gesteld aan “gemeten”) en hoe vaak professionaliseringsinterventies onderdeel zijn van de studies.

Beantwoording van de onderzoeksvragen

- 1) Hoe zien effectieve formatieve toetspraktijken in de klas eruit? Met andere woorden, wat DOET de docent die effectieve formatieve toetspraktijken in de klas bereikt?

Kwantitatieve analyses. Hiervoor is een overzicht gemaakt van codes die te maken hebben met docentactiviteiten in de verschillende fasen. Deze codes zijn aan tekstfragmenten in de artikelen gehangen (meestal de resultatensecties) en geven een beschrijving van wat docenten wel of juist niet doen (codes “NOT phase....”). Per artikel was er dus geen limiet hoe vaak een code gebruikt kon worden (bijvoorbeeld als er vier keer in de resultatensectie staat dat de docent heldere doelen naar de studenten communiceert dat wordt dit vier keer gecodeerd met “Phase 1, explicate learning goals”). Voor de kwantitatieve analyse is het aantal codes per fase geteld. De subcodes per fase zijn niet apart geteld, omdat dit een vertekend beeld kan geven doordat deze codes ongelimiteerd per artikel zijn gebruikt.

Omdat in deze studie naar formatief toetsen gekeken wordt vanuit een procesbril bestaande uit vijf fasen, is berekend hoe vaak de verschillende fasen samen in een studie aan bod kwamen. Hiervoor is de ‘co-occurrence’ berekend in Atlas-ti.

Kwalitatieve analyse. Per fase zijn alle gecodeerde fragmenten onder elkaar gezet, inclusief alle sub-code fragmenten. Deze fragmenten zijn door beide onderzoekers onafhankelijk van elkaar gelezen om thema’s te identificeren die kenmerkend waren voor deze fase, zowel gericht op wat de docent juist *wel doet* als op wat de docent *niet doet*. De geïdentificeerde thema’s zijn naast elkaar gelegd om overeenstemming of verschillen te duiden. Na overleg is een overeengekomen set aan thema’s benoemd die een fase kenmerken in termen van wat “wat de docent doet” en “wat de docent *niet* doet”. Met deze overeengekomen lijst, zijn alle fragmenten nogmaals bekeken om concrete voorbeelden behorende bij de geïdentificeerde thema’s te selecteren.

- 2) Onderzoeksvraag 2: Welke effect op studenten hebben deze toetspraktijken in termen van leeractiviteiten, leerprocessen en of leeruitkomsten?

Voor het beantwoorden van onderzoeksvraag 2 is eerst geteld hoe vaak de code “student learning outcome measured” is gegeven (code gehangen aan de titel van elk artikel, dus maximaal 1 keer per artikel gebruikt). Dit aantal geeft aan in hoeveel van de geanalyseerde artikelen een leeruitkomst is gemeten (onafhankelijk van het type leeruitkomst of de gehanteerde onderzoeksmethode). Daarna zijn in deze selectie van artikelen alle fragmenten van de codes “measured learning outcome” en “teacher behaviour → effect on student” onder elkaar gezet. Deze fragmenten zijn in de artikelen gebruikt om te duiden wat voor leeruitkomst gemeten wordt (“measured learning outcome”) of welke leeruitkomst gestimuleerd wordt door bepaald docent gedrag (“teacher behaviour → effect on student”). De resulterende fragmenten zijn op twee manieren kwalitatief gecodeerd. Ten eerste, door beide onderzoekers is uit de lijst met alle fragmenten een set van gemeten leeruitkomsten/leerprocessen geselecteerd, gecategoriseerd binnen een aantal “typen” leeruitkomsten. Deze analyse geeft dus een overzicht van welke “typen” leeruitkomsten gemeten worden in de studies en hoe vaak iedere type leeruitkomst voorkomt. Ten tweede is vanuit de code “teacher behaviour → effect on student” een overzicht gemaakt van welke docentgedragingen resulteerden in welke leeruitkomst. Hierbij zijn alle docentgedragingen gekoppeld aan een of meerdere fasen van de FT-cyclus. Dit overzicht biedt een heel concreet overzicht van voorbeelden van docentgedrag binnen bepaalde FT-fasen en het specifieke effecten hiervan op studenten.

- 3) Onderzoeksvraag 3: Welke formatieve toetscompetenties heeft een docent nodig om deze effectieve formatieve toetspraktijken in de klas te kunnen vormgeven?

Om antwoord te geven op de vraag wat docenten moeten kennen en kunnen om formatieve toetspraktijken in de klas te realiseren, zijn de daadwerkelijk gerealiseerde toetspraktijken (zie onderzoeksvraag 1) als uitgangspunt genomen voor het destilleren van noodzakelijke docentcompetenties. Er is dus geen expliciete zoekstrategie uitgezet op docentcompetenties voor formatief toetsen an sich, los van de praktijken die ze daarmee bereiken. Voor het beantwoorden van onderzoeksvraag 3 zijn twee strategieën gebruikt: 1) tijdens de coderingsfase zijn tekstfragmenten uit de artikelen die refereerden naar wat de docent moet kennen en kunnen om de FT gedragingen (uit onderzoeksvraag 1) effectief uit te voeren gecodeerd als “teacher competencies for formative assessment”. Kwantitatief is gekeken naar hoe vaak de code gegeven is, verdeeld over hoeveel artikelen. Kwalitatief zijn al deze tekstfragmenten zijn onder elkaar gezet en door beide onderzoekers onafhankelijk van elkaar gecomprimeerd tot een set van noodzakelijke competenties. 2) Nadat onderzoeksvraag 1 beantwoord was, en we dus een overzicht hadden van de gedragingen die het verschil maken tussen een meer of minder effectieve docent, hebben beide onderzoekers eerst los van elkaar en daarna gezamenlijk een overzicht gemaakt van noodzakelijke competenties die voortkwamen uit deze effectieve docentgedragingen in de vijf FT-fasen. Uit strategie 1 en 2 is een totaaloverzicht gemaakt van noodzakelijke formatieve toetscompetenties en zijn deze gekoppeld aan een of meer fasen van de formatieve toetscyclus en ondersteund met voorbeelden uit de tekstfragmenten.

- 4) Onderzoeksvraag 4: Aan welke formatieve toetscompetenties (ozv3) en/of gedragingen (ozv1) besteden professionaliseringstrajecten aandacht? Op welke gebieden zijn deze succesvol en waarop (nog) niet?

Per artikel is gecodeerd of er sprake was van een professionaliseringsinterventie (ja/nee), hoe deze professionalisering eruit zag qua inhoud en vorm, op welke fasen van de FT-cyclus deze aanspraak deed (“description of professionalisation approach”). Vervolgens zijn de effecten van deze interventies op docenten gecodeerd (“effect of professionalisation on teacher behaviour”) en gekoppeld aan de fasen van de FT-cyclus. Deze analyse gaf inzicht in de fasen van de FT-cyclus die aangesproken worden in professionaliseringsinterventies en op welke van deze fasen de betreffende interventies daadwerkelijk tot ander docent gedrag leidden of waar dit gewenste/bedoelde effect niet bereikt werd.

Op basis van de resultaten op onderzoeksvraag 1 tot en met 4 zullen aanbevelingen gedaan worden voor toekomstige professionalisering.

Hoofdstuk 4. Resultaten

Algemene gegevens

Hieronder staat in cijfers aangegeven waar de geselecteerde studies plaatsvonden en over gingen. Opvallende trends hierin waren dat veel studies waren uitgevoerd in Amerikaanse context (VS; 59%) en voor het overgrote merendeel in het primair (44%) en voortgezet onderwijs (51%). De vakgebieden science, rekenen/wiskunde en taal goed vertegenwoordigd waren, met in 31 studies andere vakgebieden als bijvoorbeeld biologie, gym, psychologie, recht, of geschiedenis. Qua methodologie was 16% van de studies alleen kwantitatief, terwijl 54% alleen kwalitatief was en 30% een mix van beide.

Onderwijsniveau*

Basis Onderwijs	47
Middelbaar Onderwijs	54
Hoger Onderwijs	13
Lerarenopleidingen	9

*het totaal telt niet op tot 106, omdat sommige studies over meerdere niveaus gingen

Vakgebied*

Science	43
Talen	20
Rekenen/wiskunde	15
Anders	31

*het totaal telt niet op tot 106, omdat sommige studies over meerdere vakgebieden gingen

Land/werelddeel

USA	63
UK	9
Europa, niet UK	10
Azie	9
Anders	15

Onderzoeksmethodologie

Kwantitatief	17
Kwalitatief	57
Gemixt	32

In 42 van de 106 studies wordt een student leeruitkomst gemeten ("Student learning outcome measured"). Dit is in bijna 40% van de studies. Zowel de leeruitkomsten als de metingen variëren hierin sterk, waarbij geen (streng) eisen zijn gesteld aan de degelijkheid van de metingen. Dit kan bijvoorbeeld variëren van de docent heeft in de klas ervaren dat de leerlingen meer toegewijd zijn, tot scores op landelijke toetsen rekenen, tot studentvragenlijsten over motivatie. De resultaten bij onderzoeksvraag 2 en Appendix 2 gaan uitgebreider op deze leeruitkomsten in.

In 43 van de 106 studies beschrijft de studie een professionaliseringstraject rondom formatief toetsen als interventie. Dat is in zo'n 41% van de studies. Deze studies zijn uitgebreid kwalitatief geanalyseerd voor onderzoeksvraag 4.

Onderzoeksvraag 1

- 1) Hoe zien effectieve formatieve toetspraktijken in de klas eruit? Met andere woorden, wat DOET de docent die effectieve formatieve toetspraktijken in de klas bereikt?

De kwantitatieve resultaten

De kwantitatieve gegevens laten het volgende zien patroon zien (zie Tabel 3 en 4). Fase 2, gericht op het ontlocken en verzamelen van studentreacties, krijgt beduidend de meeste aandacht in de bestudeerde studies (32,6%). Toetsen kan natuurlijk ook niet bestaan zonder dat er studentdata verzameld worden, dus dit is wellicht de meest voor de hand liggende stap van de Formatieve Toets-cyclus (FT-cyclus). Fase 3 en 5, respectievelijk het analyseren en interpreteren van studentdata (fase 3) en het aanpassen van instructie of leergedrag (fase 5) worden beide in 22% van de gevallen aangesproken. Opvallend is dat fase 4, feedback geven, slechts in 9% van de fragmenten aangesproken wordt, terwijl dit altijd wordt gezien als de kern van formatief toetsen. Fase 1, het benoemen van leerdoelen en succescriteria, krijgt met 14,5% ook beduidend minder aandacht dan fase 2, 3 en 5. Tabel 5 laat zien in hoeverre de bestudeerde artikelen de verschillende fasen van de formatieve toetscyclus samen bestudeerde (co-occurrence). Hierin is te zien dat fase 1 en 4 vaak samen voorkwamen en fasen 2-3, 2-5 en 3-5. Ook de kwalitatieve analyses lieten duidelijk een trend zien van artikelen die aandacht besteden aan de fasen 2-3-5 samen (zie hieronder).

Tabel 4 geeft ook informatie over gedrag dat docenten *niet* lieten zien, terwijl dit door betreffende studies wel expliciet *bedoeld* was (de "NOT" coderingen). Hier scoren fase 5 en fase 2 het hoogste. Deze resultaten suggereren dat docenten het minst slagen in het toepassen van fase 2 en fase 5 gedrag. De kwalitatieve analyses geven meer inzicht in wat hier daadwerkelijk gebeurt.

Tabel 3. Mate waarin de fasen van de FT-cyclus voorkomen.

	Fase 1	Fase 2	Fase 3	Fase 4	Fase 5
Totaal aantal	163	376	254	104	255
% *	14,1%	32,6%	22%	9%	22%

* van totaal aantal fase coderingen

Tabel 4. Mate waarin de bedoelde fasen van de FT-cyclus NIET worden uitgevoerd.

	NOT fase 1	NOT fase 2	NOT fase 3	NOT fase 4	NOT fase 5
Totaal aantal	49	77	61	41	79
% *	16%	25%	20%	13%	26%

* van totaal aantal NOT fase coderingen

Tabel 5. Co-occurrence tabel van de vijf fasen

	Fase 1	Fase 2	Fase 3	Fase 4	Fase 5
Fase 1	-	20%	14%	36%	16%
Fase 2	-	-	31%	26%	38%
Fase 3	-	-	-	11%	55%
Fase 4	-	-	-	-	19%

De kwalitatieve resultaten

Veel van de bestudeerde studies tonen voorbeelden van docenten die minder goede, goede en nog betere formatieve toetspraktijken realiseren in de klas (b.v. Lee, 2011; Penuel, Boscardin, Masyn, & Crawford, 2007; Ruiz-Primo & Furtak, 2006; 2007; Wylie & Lyon, 2015) of laten zien welke belangrijke, succesvolle, veranderingen docenten aanbrengen in hun formatieve toetspraktijken in de klas (bv. James & Moreland, 2005; Pham & Renshaw, 2015; Torrance & Pryor, 2001). Deze paragraaf beschrijft eerst typerend docent gedrag per fase. Hierbij gaat het om kritische gedragingen die het onderscheid maakten tussen docenten die minder goede, goede, en betere formatieve toetspraktijken in de klas realiseerden..Hierbij gaat het niet per definitie om docenten die meer of minder effectief zijn in termen van “het hebben van effect op student leeruitkomsten”, omdat deze effectmeting slechts in een klein deel van de artikelen terugkomt (zie onderzoeksvraag 2). In Appendix 2 is een overzicht te vinden van gevonden effectief docentgedrag in termen van effect op leeruitkomsten. Bij het beantwoorden van onderzoeksvraag 1 noemen we docentgedrag ook effectiever als de docenten gedrag laten zien dat meer voldoet aan wat theoretisch, in het betreffende artikel, wordt gezien als goed formatief toetsen. Zo beschrijven bijvoorbeeld veel artikelen casussen van docentactiviteiten of docent-student interacties en beschrijven hierin voorbeelden van docenten die meer of minder aan hun theoretisch kader voor formatief toetsen voldoen (bv. Ruiz-Primo & Furtak, 2006;2007 “de ESRU-cyclus”; van der Pol et al., 2012; 2013 “de cyclus van contingent teaching”). Andere studies beschrijven professionaliseringsinterventies waarbij de effectievere docent de docent is die na afloop gedrag laat zien dat door de interventie getriggerd is (bv. Sabel, Forbes, & Zangori, 2015; Sato, Wei, & Darling-Hammond, 2008). Weer andere studies beschrijven (quasi-)experimentele studies waarbij de effectievere docent gedrag laat zien dat door de interventie bedoeld is (bv. Panadero, Tapia, & Huertas, 2012). Deze verschillende typen studies geven informatie over typerend gedrag van de docent in de vijf fasen van de formatieve toetscyclus.

Als laatste beschrijft deze paragraaf een gevonden trend die zichtbaar werd bij het analyseren van de 106 artikelen door de bril van de formatieve toetscyclus. Deze trend strekt verder dan één fase, maar beschrijft bepaald docentgedrag over een aantal fasen heen laten zien, namelijk in de kleine cyclus van fase 2-3-5.

In onderstaande beschrijvingen van de fasen staat, omwille van de leesbaarheid, steeds een selectie van referenties genoemd en staat een aantal sprekende voorbeelden van concreet docentgedrag in Boxen beschreven. Aanvullende voorbeelden zijn te vinden in Appendix 2. Hierin staat een uitgebreide tabel met beschrijvingen van concreet docentgedrag gekoppeld aan bepaalde fasen van de formatieve toetscyclus en resulterend in concrete studentuitkomsten (onderzoeksvraag 2). In onderstaande beschrijvingen zal dus geregeld verwezen worden naar voorbeelden uit deze tabel in Appendix 2.

Typerende docentgedragingen in fase 1: Verwachtingen verhelderen

Fase 1 gaat over het verhelderen van verwachtingen. Dit heeft betrekking op leerdoelen en succescriteria. De studies in deze overzichtsstudie laten zien dat docenten dit op verschillende manieren doen. Hierin zijn vier thema's te onderscheiden:

1. Formuleren van heldere en doelgerichte leerdoelen, succescriteria en learning progressions als voorwaarde voor formatief toetsen. Dit houdt ook in dat de doelen mede bepaald worden door het doel dat de docent met formatief toetsen heeft
2. Communiceren van leerdoelen en succescriteria naar studenten, op meerdere manieren en momenten
3. Actief betrekken van studenten en gezamenlijk expliciteren
4. Balans tussen helder gespecificeerde doelen en te dichtgetimmerde doelen

Ten eerste is het helder en doelgericht formuleren van leerdoelen en succescriteria voorwaardelijk voor het kunnen communiceren van deze doelen aan / met de studenten. Wylie en Lyon (2015) beschrijven dat docenten leerdoelen beschrijven in termen van *wat bereikt moet worden* komen tot een robuustere implementatie van formatief toetsen dan docenten die leerdoelen formuleren in termen van *activiteiten die uitgevoerd moeten worden*. Dat is eigenlijk het verschil tussen leerdoelen formuleren over het WAT in plaats van over het HOE (Baartman & Gulikers, 2014). Verschillende studies rapporteren op basis van observaties van docenten en/of analyses van docentreflecties dat veel docenten geen leerdoelen hebben, deze niet communiceren en/of dat de kwaliteit ervan laag is (e.g. Moss, Brookhart, & Long, 2013; Barnhart & Van Es, 2015). Chroinin and Cosgrave (2013) beschrijven een concrete en succesvolle strategie waarmee docenten hun leerdoelen en succescriteria helder kunnen krijgen: de WALF (“What Are we Learning For”) en WILF (“What I am Looking For”) strategie. Docenten die zich voorafgaand aan hun onderwijs deze vragen stellen, kunnen beter richting studenten communiceren wat zij van hen verwachten. Gearhart en collega’s (2006) en Gearhart en Osmundson (2008) beschrijven hoe docenten leerdoelen formuleren die niet één correct idee of antwoord representeren, maar zijn gericht op de onderliggende belangrijke concept of kernideeën. Die leerdoelen worden bijvoorbeeld weergegeven in een concept map in plaats van een lijst, zodat ook onderliggende relaties zichtbaar worden.

Naast het belang van inhoudelijke leerdoelen, laten verschillende studies het belang van “learning progressions” zien (bv. Falk, 2012; Furtak & Heredia, 2014; Gearhart et al., 2006). In learning progressions zie je de manier waarop studenten hun kennis en begrip in dit vak ontwikkelen, De learning progressions in de artikelen in deze review zijn vaak gericht op wiskunde, aardrijkskunde, natuurkunde of biologie (Furtak & Heredia, 2014; Furtak, 2012) en een enkele keer op taal (Bailey et al., 2014). Docenten die goed zicht en grip hebben op de learning progressions, bereiken betere formatieve toetspraktijken in de klas. De docent heeft dan zicht op hoe studenten in dit vak kennis en begrip ontwikkelen, maar vooral ook op de misconcepties die studenten rondom het thema hebben. Het zicht hebben op de learning progression is echter niet alleen van belang voor fase 1, maar lijkt een voorwaarde om als docent überhaupt tot goede formatieve toetspraktijken te kunnen komen en zal dan ook terugkomen in onderzoeksvraag 3 over wat de docent moet kennen en kunnen om goede formatieve toetspraktijken vorm te geven.

Docenten die sterk gericht zijn op het stimuleren van autonomie, verantwoordelijkheid voor en sturing door de student hebben leerdoelen die niet alleen gericht zijn op inhoud maar vooral ook participatie, sociale norm en/of samenwerking (Davis, 2011; Hawe & Dixon, 2014; Pham&Renshaw_2015; Torrance & Pryor, 2001; Willis, 2011). Dat wil zeggen dat de leerdoelen expliciet aandacht besteden aan de rol van de student en de manier waarop studenten met elkaar en met de docent om horen te gaan. Deze studies laten zien dat dit soort leerdoelen ertoe leiden dat de student-docent relaties en de student activiteit en participatie in de klas daadwerkelijk uitgedaagd wordt en leidt tot actieve studenten en meer gelijkwaardige docent-student relaties die nodig lijken te zijn voor goede formatieve toetspraktijken.

Ten tweede verhelderen docenten de leerdoelen en succescriteria op meerdere manieren en op meerdere momenten gedurende de lessen en pakken hier tijdens een les of lessenreeks herhaaldelijk op terug (bv. Moss et al., 2013; Parr & Limbrick, 2010; Schneider & Andrade, 2013). De hoogvliegers in de studie van Moss et al onderscheiden zich bijvoorbeeld doordat zij leerdoelen én succescriteria delen met studenten via rubrics, voorbeeldproducten van studenten en het voordoen van hoe het moet (modelling). Succesvolle docenten pakken bovendien aan het eind van hun lessen steeds even terug op de leerdoelen door met studenten te evalueren wat ze nu geleerd hebben met betrekking tot de leerdoelen (bv. Conway, 2011, zie Box 2; Wiliam et al. 2004). Een concreet voorbeeld hoe docenten op verschillende manieren de doelen naar studenten communiceren, is beschreven door Kang en collega’s (2016). Zij beschrijven hoe docenten verschillende scaffolding tools ontwikkelen die de studenten sturen op de juiste concepten in de lessen. Zo gebruiken docenten “word boxes” met belangrijke begrippen, rubrics met kwalitatieve beschrijvingen van hoe een goed antwoord eruit ziet, en “sentence frames”, startzinnen die aangeven wat er wordt verwacht bij een antwoord

(bijvoorbeeld een model, of een analyse). Studenten gebruiken deze word boxes, rubrics en startzinnen bij het beantwoorden van vragen (zie ook fase 2).

Ten derde laat een deel van de artikelen zien hoe docenten hun studenten actief betrekken bij het verhelderen van leerdoelen en succescriteria (Clark, 2012). Een herhaaldelijk beschreven succesvolle strategie bij het vormen van succescriteria is om voorbeelden van studentresponsen/-producten, zowel goede als minder goede voorbeelden (Fluckiger et al., 2010), en tevens authentieke producten uit 'de echte wereld', samen met studenten te bekijken om hier gezamenlijk succescriteria uit te destilleren (oa. Strowbridge, 2008; Willis, 2011). Een leuk voorbeeld wordt beschreven door Hawe en Dixon (2014) voor basisschoolstudenten in schrijfonderwijs (zie Box 1). Echter, hoewel in de bestudeerde studies vaak benoemd wordt dat succesvolle docenten studenten actief betrekken bij het begrijpen van de leerdoelen en succescriteria, wordt slechts enkele keren concreet gemaakt hoe docenten dit daadwerkelijk doen, of blijken studenten en docenten totaal andere beelden te hebben bij de leerdoelen en succescriteria (Bloxham, 2010; De Listle, 2015; Hogan et al., 2009; Newby & Winterbottom, 2011). Deze studies concluderen allemaal dat docenten te weinig doen aan het expliciteren en helder krijgen van de doelen en succescriteria voor studenten. Of zoals Rice en Smith (2011) zeggen, dat docenten niet gewend zijn dit te doen en dit een transitie in handelen vraagt die erg moeilijk lijkt te zijn. Deze mismatch tussen studenten en docenten kan echter wel resulteren in (1) verkeerde producten / studentreacties (i.e., de student geeft een reactie die de docent eigenlijk helemaal niet bedoelt); (2) ineffectieve (peer) feedback (feedback wordt gegeven op onderdelen die eigenlijk niet relevant zijn); (3) onverwachte beoordelingen (studenten worden op andere punten beoordeeld dan ze zelf verwacht hadden): *"teachers' failure to share these tacitly held criteria with students resulted in misalignment between the evaluative judgments made by students and those made by Audrey (i.e., the teacher)"* (Hawe & Dixon, 2014). Tevens is maar een klein aantal studies gevonden waarbij studenten daadwerkelijk zelf leerdoelen moeten opstellen, deze leerdoelen mogen verschillen per student (Parr & Limbrick, 2010) en/of deze leerdoelen mogen mee ontwikkelen gedurende een vak (Hawe & Dixon, 2014; Lorente & Kirk, 2013; Kearney, 2013). :

Box 1. Samen succescriteria ontdekken

Prior to commencing the writing unit, students had been exposed to and discussed a range of cartoons and comic strips. Initially, the class brainstormed features of cartoons and comics, and how authors 'got their message across' to readers. As these ideas were progressively refined, Kate recorded them on the class whiteboard where they served as a semipermanent, living record for the duration of the unit. The teacher also provided students with an assessment rubric, that overlapped to a large extent with the student developed criteria

Hawe & Dixon, 2014, P.73

Box 2. Samen met studenten criteria en assessmentactiviteiten bedenken

At the beginning of the unit both I and my students discuss the pathology of assessment in education (Santos Guerra, 1988), focusing on current problems and reflecting with students on their experiences in the different subjects throughout their compulsory schooling and non-compulsory education too, but with a special focus on their experiences in physical education. Afterwards, they discuss the unit's assessment by thinking through the following question: what can we do to avoid similar issues when assessing our learning in the unit? Students, in groups of three or four, organise different proposals which are shared with the whole class later on to reach an agreement. The only condition is that assessment activities must assess the unit's objectives that include, among others, designing physical education assessment tasks.

Lorente & Kirk, 2013, p. 83

Verschillende studies laten ook het belang zien van het vinden van een balans tussen heldere, transparante criteria en niet te dichtgetimmerde criteria (Aschbacher & Alonzo, 2006; Hawe & Dixon, 2014; Torrance, 2004). Wanneer criteria helder en zeer concreet ontstaat er "criterion compliance" (Hawe & Dixon, 2014; Torrance, 2007; 2004) waarbij het leerproces van studenten puur gericht is op het voldoen aan criteria. Zeker wanneer de

docent met formatief toetsen andere doelen nastreeft dan “uit het hoofd leren”, zoals bijvoorbeeld het stimuleren van zelf-regulatie, werkt het hebben van een te concreet lijstje met criteria niet goed. Gearhart et al. (2006) en Gearhart & Osmundson (2008) laten zien dat criteria niet moeten bestaan uit een lijst concepten, maar gericht moeten zijn op de onderliggende verbanden en kern ideeën. Succesvolle docenten representeren leerdoelen bijvoorbeeld in een concept map in plaats van in een lijst met concepten. Hawe & Dixon (2014), Lorente & Kirk (2013) en een vergaand voorbeeld van Conway (2011) laten zien dat succesvolle docenten in staat zijn om kaders te stellen voor leerdoelen (over welke thema’s moeten de leerdoelen gaan), waarbinnen studenten zelf vrijheid krijgen om deze concreet te maken passend bij hun eigen niveau of interessegebied. Twee concrete voorbeelden van hoe dit eruit kan zien zijn te lezen in Box 3.

Box 3. Voorbeeld van docent die gezamenlijk met de klas leerdoelen opstelt, verheldert en steeds weer terugpakt

To help them with this stage of the enquiry we use a fantastic set of ‘Thinking Dice’ that provide appropriate forms of phrasing to stimulate questions linked to the different types of thinking categorised in ‘Bloom’s Taxonomy’. One of the great advantages of these dice is that they allow all students to engage with, and participate in, the process of framing questions. Giving lower attaining students the ‘Know’ and ‘Understand’ dice enables them to formulate questions from starters such as ‘What is meant by...?’ and ‘Why did...?’. Higher attainers can work with the ‘Analysing’ and ‘Evaluating’ dice, using stems such as ‘What is the relationship between ... and...?’ and ‘What was the most important...?’

My role as a teacher is to help the group leaders to phrase and record the questions they generate, and to type them into a central document as the lesson progresses. By the end of the activity we have an extensive list of questions that the class feel merit further work. At this stage, I include all suggestions, even repetitive or poor questions, and I can quickly print out the document to distribute to the groups. They then take responsibility for eliminating repeats and selecting the best (in terms of depth and scope) from those that are similar. I act as a kind of facilitator and secretary, ensuring that the central copy, which is projected on to the whiteboard, represents our collective decisions as different groups put forward their ideas to be discussed and accepted, refined or rejected as a class.

Each lesson we put aside at least ten minutes for evaluation, asking four questions. The first is ‘What have we learned?’ On the wall of my classroom for the whole term is a large-scale printed list of the original questions generated by each class, grouped into lessons. Each student has a copy in their file.

Conway (2011, p.53)

The second example describes a robust implementation of learning expectations where the teacher involved students in the development of the actual learning intention and focused on what students would know or understand by the end of the lesson:

I provide students the main topic for this section and they use whiteboards to develop questions or statements that will become the main learning intentions for the current subject. This helps students understand what we are trying to learn about in class and gives them some ownership in the process ... We were studying hurricanes and one of the students came up with the question of why do hurricanes spin. As a class, with guidance, we refined the question to ‘What factors cause a counter-clockwise rotation of hurricanes in the northern hemisphere?’ ... As the teacher I give students the main subject to develop learning intentions around and then help guide the revision of the question when it is presented ... I revised which areas I focused my lesson on

Wylie & Lyon (2015, p 150). Robuuste implementatie van de fase “clarifying learning goals”

Typerende docentgedragingen bij fase 2: Ontlokken en verzamelen van studentreacties.

Onder fase 2 vallen alle mogelijke manieren, methodieken en strategieën die docenten gebruiken om data (i.e. reacties, responsen) van studenten te ontlokken en te verzamelen. De thema’s die hierin terugkwamen, zijn:

1. Ontwerpen en inzetten van doelgerichte methodieken, gekoppeld aan leerdoelen
2. Het repertoire aan methodieken, formeel en informeel, dat de docent gebruikt in de klas
3. Gebruik maken van vragen stellen en klasdiscussies gericht op het uitdiepen van begrip in plaats van zoeken naar het goede antwoord
4. Mate waarin de discussies meer studentgestuurd zijn

Een eerste thema dat in deze fase opvalt is dat de effectievere docenten in staat zijn om methodieken doelgericht in te zetten, dat wil zeggen rechtstreeks gekoppeld aan de leerdoelen. De docent stelt doelgericht vragen en ontwerpt doelgerichte activiteiten die studentreacties ontlokken die direct gekoppeld kunnen worden aan de mate waarin de studenten aan de leerdoelen en succescriteria voldoen (bv. Jones & Moreland, 2005; Falk, 2012). Andere studies laten ook zien dat docenten die minder goede formatieve toetspraktijken bereiken, niet in staat zijn om assessment taken, prompts of vragen te ontwerpen die daadwerkelijk informatie over de leerdoelen ontlokken (bv. Aschbacher et al., 2006; Beesley, 2009). Moss en collega's (2013) laten zelfs zien dat wanneer docenten hun leerdoelen niet helder hebben, 50% van hun leer- en toetsactiviteiten ongereleerd waren aan leerdoelen.

Furtak & Heredia (2014) beschrijven hoe learning progressions (zie fase 1) docenten konden helpen bij het doelgericht vormgeven van formatieve toetsmethodieken. De docenten in deze studie ontwikkelden bijvoorbeeld scenario/casusvragen die studenten moesten interpreteren, en mc-toetsen waarin ze de veel voorkomende misconcepties verwerkten als antwoordalternatieven. Naast de inhoud van de methodieken, gaat het ook om de timing ervan. De methodieken moeten passen bij de volgorde van de instructieactiviteiten in de lessenserie (Falk, 2012), en moeten worden ingezet op logische momenten in een lessenserie, bijvoorbeeld op een moment waarop een docent wil starten met een nieuw / complexer onderwerp waarvoor het belangrijk is om te inventariseren of de studenten het voorgaande / voorwaardelijke goed hebben begrepen (Ayala et al., 2008).

Een tweede kenmerk is dat effectieve docenten een groter repertoire hebben van methodieken, zowel formele als informele, om data van studenten te ontlokken en in hun dagelijkse praktijk ook gebruik maken van deze variatie (Herman, Osmundson, Dai, Ringstaff & Timms, 2015; Sato, 2008; Suurtamm, Koch & Arden, 2014). Deze studies laten tevens zien dat docenten die groeien in hun formatieve toetspraktijken vooral meer gebruik gaan maken van informele methodieken om in iedere dagelijkse les zicht te krijgen op het begrip van studenten. Matuk en collega's (2015) beschrijven bijvoorbeeld hoe een digitaal systeem de docenten in staat stelde om veel meer informatie te achterhalen over het begrip van individuele studenten. De docenten vergeleken de scores van de studenten in het systeem, maar gebruikten bijvoorbeeld ook informatie over voorgaande schooljaren om te begrijpen waar bepaalde hiaten in begrip vandaan komen. Falk en collega's (2007) en Bailey en collega's (2014) beschrijven allebei hoe docenten op basis van een rubric of een learning progression de studenten observeren in de klas, conversaties volgen en schriftelijk werk bekijken. De docenten in deze studies zijn – na ervaring te hebben opgedaan – van mening dat dit veel meer waardevolle informatie oplevert over de student dan de (standaard) test alleen. De minder effectieve docenten blijven daarentegen veelal hangen in het gebruik van methodieken die gericht zijn op het ontlokken van goed meetbare uitkomsten als feitelijke, inhoudelijke vakkennis (Davis & Neitzel, 2011).

Een derde geïdentificeerd kenmerk van fase 2 zijn twee veelvuldig gecodeerde methodieken die het verschil leken te maken tussen docenten die in de klas minder goede, goede en betere formatieve toetspraktijken lieten zien, waren vaardig vragen stellen (i.e., “questionning”) en het voeren van klas/groepsdiscussies (“whole class discussions”). Belangrijk hierbij was dat deze zijn gericht op het expliciteren en verdiepen van *begrip en misconcepties* in plaats van op het verkrijgen van een goed antwoord (Aschbacher & Alonzo, 2006; Lysaght & O’Leary, 2013; Offerdahl & Tomanek, 2011; Ruiz-Primo & Furtak, 2006; 2007; Sato et al., 2008; Suurman et al., 2010; Torrance & Pryor, 2001). Aschbacher & Alonzo (2006) bestudeerden hoe docenten notebooks van studenten gebruikten voor formatief toetsen. Zij identificeerden uit de rapportages en observaties van acht docenten die zij volgden vier manieren van vragen stellen, waarbij de vierde de meeste mogelijkheden biedt voor formatief toetsen (zie Box 3). Deze manier werd echter slechts door één docent tot uitvoering gebracht. Ook andere studies lieten zien dat docenten moeite hadden met het vormgeven van assessment taken, vragen en discussies die gericht zijn op conceptueel begrip in plaats van te zoeken naar een goed antwoord op meer feitelijke kennis (Ateh, 2015; Gomez & Jakobsson, 2014; Offerdahl & Tomanek, 2011; Riggan & Nabors Olah, 2011). Davis & Neitzel (2011) keken in het bijzonder naar formatieve toetspraktijken gericht op het stimuleren

van zelf-regulerend leren en ontdekten dat docenten het niet goed voor elkaar kregen om leerdoelen en assessment taken te ontwikkelen die gericht waren op zelf-regulerend leren. Ook deze docenten bleven gericht op feitelijke, makkelijk te meten kennis.

Box 4. Vier manieren van vragen stellen

Our examination of how the teacher used notebooks revealed four typical patterns of use, somewhat reminiscent of the Goldilocks story—too little, too much, and just right:

- 1. minimal guidance to students in what to write, which is often very vague and seldom conceptually focused. Eg “Write what you learned today”;*
- 2. low guidance that provides a little structure but still insufficient focus;*
- 3. overly prescriptive guidance that promotes student copying of “the right” information rather than expressing their own ideas; and*
- 4. moderate guidance, the “just right” type of questioning and directions that provide conceptual focus, allow students to do their own thinking related to the key unit concepts, and encourage an accurate and complete record of data and thinking in the notebook.*

Aschbacher & Alonzo, 2006, p.190

Barnhart en Van Es (2015) beschrijven bijvoorbeeld hoe een docent dóórvraagt op de achterliggende ideeën van studenten, om te achterhalen hoe diep hun begrip is, in plaats van snel tevreden te zijn met een (algemeen) antwoord. Gearhart en collega's (2006) beschrijven een voorbeeld van een docent die groep discussies gebruikt om een beeld te krijgen van het begrip van studenten (zie Box 5)

Box 5: voorbeeld van het inzetten van groepsdiscussies om dieper begrip boven tafel te krijgen

Yvette organized her students into small groups for a follow-up activity. She assigned each group one assessment question, and asked students to record their response on a whiteboard; as each group presented their work to the class, Yvette engaged the students in short discussion. This informal assessment had greater impact on her practice than any examination of the students' written responses; in her portfolio, Yvette wrote that there was “adequate evidence of students' understanding of the concept” because “the class seemed to have at least a ninety percent acquisition” based on whiteboards (YJ, Fall 2003 portfolio).

Gearhart et al, 2006, p.246

Een aanvullende cruciale factor die hierbij het verschil maakt tussen docenten die meer of minder succesvolle formatieve toetspraktijken realiseren, is of de docent erin slaagt de klasdiscussies studentgestuurd te maken (e.g. Lysaght & O'Leary, 2013; Penuel, 2007; Ruiz-Primo & Furtak 2006; 2007). Dat wil zeggen, de docent stelt op zo'n manier vragen en geeft bewust activiteiten vorm die studenten zelf en onderling 'dwingen' om hun begrip en misconcepties te expliciteren en uit te diepen (bv. Wylie & Lyon, 2015). Een mooi voorbeeld komt uit Pham en Renshaw (2015) die twee casussen van docenten beschrijven in een hoger onderwijs context in Vietnam, in een cultuur gekenmerkt door sterk docentgestuurd en traditioneel onderwijs. Beide docenten experimenteren met mogelijkheden om meer formatieve toetspraktijken te realiseren. De grootste impact en verandering vindt plaats bij de docent Recht die in staat is om via een verplichte “vragen-stel-strategie” de student-docent en student-student relaties in de klas fundamenteel te veranderen. Met positieve leereffecten als resultaat (zie onderzoeksvraag 2 en Appendix 2). Ondanks dat docenten die groeien in hun formatieve toetspraktijken meer gaan doorvragen en in sterk toenemende mate aan hele klasdiscussies doen, blijven veel praktijken docentgestuurd. Het echt veranderen van de docent-student relaties, waarbij de student en peers de vragen stellen en de invulling van het assessment bepalen, gebeurt maar zeer mondjes maat (o.a. Lysaght & O'Leary, 2013). Van der Pol, Volman, Oort & Beishuizen (2013) en Feldman & Capobianco (2008) voegen hieraan toe dat de meest succesvolle docent in staat is om de manier van vragen stellen en de sturing hierin aan te passen aan het begripsniveau van de student:

He (i.e., successful teacher) used what the students came up with but also steered the discussion when he saw that their understanding was not deep enough. The teacher and students deepened their understanding of the concept of community together. (van der Pol et al., 2013, p. 633)

Docenten die nog geen doorleefd begrip hadden van formatief toetsen (scaffolding in deze studie) stelden meer sturende vragen aan studenten, omdat zij dachten dat dit niet mocht in formatief toetsen. Zij hadden een misconceptie over “prompten” die hen beperkte in het vormgeven van effectieve formatieve praktijken voor alle studenten.

Box 6 laat het verschil zien tussen drie docenten die een goede (docent 1) en een nog betere informele formatieve toetspraktijk laten zien (docent 2 en 3). Beide stellen vragen en hebben klasdiscussies die gericht zijn op begrip en het uitdiepen van dit begrip, maar de eerste docent is meer docentgestuurd terwijl de tweede docent de discussies veel meer laat leiden door de studenten (Ruiz-Primo & Furtak, 2006).

Box 6: een gedeeltelijk succesvolle en twee succesvolle docenten in fase 2

Docent 1: Diana. Gedeeltelijk effectieve docent (Ruiz-Primo & Furtak, 2006, p. 219)

She poses a question, a student responds, and then Diana repeats or revoices the comment, elaborates on the comment, and provides some kind of evaluative response. Repeating this sequence over and over during a discussion creates a conversation characterized by a good deal of guidance on the part of the teacher, as she asks questions that are easily answered by students, then she praises the students and adds more information. Occasionally, Diana will push a student to think more about her response, or to provide more information, but then will return to the pattern. In this way, Diana controls the flow of conversation and clearly lets students know if they have the answer for which she is looking.

Docent 2: Carol. Meest effectieve docent (Ruiz-Primo & Furtak, 2006, p. 223)

An analysis of Carol's teaching revealed the presence of informal assessment practices during all discussions. She was the teacher who used the most diverse strategies across all the ESRU phases; for example, she was the teacher who focused the most on cycles linked to the conceptual domain. She frequently used the strategy of responding to a student's contribution by asking another student to repeat it. This strategy helped to make a key idea more explicit in the discussion at hand. Carol was also the teacher with the most student-initiated cycles.

She (= Carol) held discussions that were longer and more frequent than those of the other teachers. Carol frequently began a discussion by reviewing what had been addressed the previous day, and then posed open-ended questions to the class. She would frequently follow the course of students' comments rather than pursuing her own narrative for the class.

Docent 3: Danielle. Meest effectieve docent (Ruiz-Primo & Furtak, 2007, p. 72)

In contrast to Rob, Danielle starts her conversation in this excerpt with student input, and builds upon it by repeating students' words (accepting them in to the ongoing conversation), asking students to elaborate, and clarifying and elaborating upon student comments. This pattern occurred many times in our observations of Danielle's teaching. This example thus fits the model of an assessment conversation, in that Danielle's questions are more instructionally responsive than would be expected from an everyday lesson. That is, Danielle asks questions similar to those Rob asks his students, and also frequently repeats what they have said; however, Danielle pushes her students for more elaboration on their responses, and also provides feedback that allows students to learn more about classroom expectations for learning. Danielle's lessons also showed characteristics of assessment conversations not found in Rob's or Alex's lessons. For example, she was the only one comparing and contrasting students' responses. Rather than moving from student to student without identifying how students' ideas were different.

NB. In de artikelen van Ruiz-Primo & Furtak (2006; 2007) staan uitgewerkte voorbeelden van de assessment gesprekken die deze docenten met hun studenten hebben.

De voorbeelden in Box 5 laten zien dat deze meer of mindere succesvolle formatieve toetspraktijk uit meer bestaat dan alleen het op een goede manier ontlokken van studentreacties (fase 2). Ruiz-Primo & Furtak (2006; 2007; 2011) gebruiken hiervoor de ESRU-cyclus: Eliciting student response – Student response – Recognising student response – Using student response. Dit komt overeen met fase 2-3-5 van onze formatieve toetscyclus. Deze “kleine” formatieve toetscyclus kwam herhaaldelijk terug in de kwalitatieve analyses. Dit is dan ook een van de meest belangrijke geïdentificeerde trends van effectieve docentgedrag die aan het einde van deze paragraaf verder besproken wordt.

Typerende docentgedragingen in fase 3: Analyseren en interpreteren van studentreacties

In het bestuderen van de artikelen over fase 3 viel een aantal dingen op in vergelijking met artikelen over de andere fasen: Een klein aantal artikelen, uit de Verenigde Staten en in de context van science onderwijs, zoomt heel gedetailleerd in op specifiek en alleen deze fase van het analyseren van studentreacties (Talanquer, Bolger & Tomerak, 2015; Talanquer, Tomerak, & Novodvorsky, 2013), onafhankelijk van de voorgaande fasen of vervolgfases. In veel andere studies komt fase 3 aan bod naast de andere fasen (veelal fasen 2-3-5), maar fase 3 wordt in veel artikelen ook niet benoemd of niet concreet en expliciet uitgediept in termen van docentgedrag (b.v. Buck et al., 2010; Phelan et al. 2012). Het gaat hier immers ook vaak om iets wat docenten in hun hoofd doen en niet direct zichtbaar is in de lespraktijk. Toch kwamen uit de kwalitatieve analyses een aantal duidelijke bevindingen naar voren die onderscheid maken tussen docenten die minder goede en betere analyse-praktijken laten zien voor deze fase van de formatieve toetscyclus. Dit zijn:

1. Analyses gericht op dieper begrip
2. Aandacht voor misconcepties
3. Weinig tijd besteed aan analyse
4. Focus op zwakheden
5. Hoeveelheid en soort informatie waarop docent conclusie baseert
6. Gebruik van technologie om analyse te ondersteunen

De gedetailleerde studies over hoe docenten studentresponses analyseren (Gearhart & Osmundson, 2008; Sabel, Forbes, & Zangori, 2015; Talanquer et al. 2013; 2015) laten zien dat docenten geneigd zijn in eerste instantie naar oppervlakte kenmerken te kijken, zoals de verzorging of schrijfstijl in plaats van naar de daadwerkelijke inhoud van de responsen. Dit vertroebelt het beeld dat de docent van studenten opdoen en kan leiden tot onterechte interpretaties en conclusies. Gearhart en Osmundson (2008) laten voorbeelden van docenten zien die beter kunnen analyseren, docenten die studentreacties heel oppervlakkig analyseren (bv. alleen kijken naar de gemiddelde score) tot aan docenten die ongepaste analysestrategieën gebruiken om conclusies te trekken. Deze oppervlakkige of foutieve manier van analyseren wordt sterk verminderd wanneer de docent (1) meer vakkennis heeft en (2) meer zicht heeft op de leerdoelen en kennisontwikkeling in het vakgebied (bijvoorbeeld in een learning progression, zie fase 1). Forbes et al (2015) beschrijven hoe betere docenten duidelijk kunnen uitleggen waarnaar zij op zoek zijn in het werk of de discussies van studenten. Deze docenten letten niet alleen op het gebruik van bepaalde begrippen of concepten, maar analyseren ook of studenten de begrippen ook daadwerkelijk goed gebruiken.

Een thema dat hierbij aansluit, en dus een onderscheid tussen minder en betere formatieve toetspraktijken laat zien, is het al dan niet gericht zijn op het analyseren van studentreacties op misconcepties (bv. Forbes et al., 2015; Matuk et al., 2015; Falk, 2012; Offerdahl & Tomanek, 2011; Talanquer et al., 2013; 2015). Vaak gebeurt dit door het gebruiken van learning progressions om de ideeën van studenten te analyseren en te categoriseren (Bailey et al., 2014; Falk, 2012; Mitton-Kukner et al., 2014). Furtak en Heredia (2014) beschrijven een voorbeeld van een docent die het werk van studenten ordent in 3 stapeltjes volgens de learning progression, terwijl deze docent eerder (voordat hij/zij een professionaliseringstraject doorlopen had) studenten alleen indeelde op goed/fout, zonder verder specificering of onderbouwing. De studie van Buck, TrauthNare en Kaftan (2010) laat zien hoeveel moeite docenten hebben met het identificeren van misconcepties in de antwoorden van studenten. Dit hield ook stand nadat docenten vooraf gesproken hadden over misconcepties in hun vak (in dit geval science). Docenten konden deze kennis vervolgens niet transfereren naar hun analyseren van studentantwoorden. Box 7 beschrijft een voorbeeld van hoe docenten “clicker and reading questions” gebruiken om zicht te krijgen op misconcepties die in de klas leven. Deze studies benadrukken tevens dat docenten dit beter kunnen indien zij meer en gedetailleerde vakkennis hebben en zicht hebben op hoe kennis in dit vakgebied ontwikkelt (zie “learning progressions” in fase 1, en onderzoeksvraag 3).

Box 7. Voorbeeld van het ontdekken van misconcepties

Mari, Judy and Brian also grew to consider clicker and reading questions as a mechanism for uncovering the specific conceptual difficulties students were experiencing. Previously, they had relied on their knowledge accumulated through years of teaching about general topics that students most often found difficult. Student reading questions in particular revealed much finer grain details of student understanding of these general topics as well as difficulties with concepts unanticipated by the instructors. Clicker questions, on the other hand, provided evidence that what the instructors were gleaning about individual student understanding in one-on-one interactions actually represented a more prominent trend within the class

Offerdahl & Tomanek, 2011, p. 789

Een heel basaal, maar herhaaldelijk terugkomende factor die docentgedrag in deze fase beïnvloedt is de tijd die docenten nemen om studentreacties te analyseren (Abrams & McMillan 2014; Herman et al. 2015; Traugh-Nare & Buck, 2011). Deze studies laten zien dat docenten, zelfs na professionalisering, maar heel kort de tijd nemen (zo'n 5 minuten maximaal) om studentantwoorden te analyseren. De analyses kenmerken zich in deze gevallen door oppervlakkige analyses vooral gebaseerd op vorm in plaats van op inhoud (zie thema 1 hierboven). Wanneer docenten echter meer tijd namen voor hun analyses en hun analyses met collega's bespraken, waren hun vervolgacties (formatieve toetscyclus fase 5) echter wel beter (Abrams & McMillan, 2014; Herman et al., 2015).

Opvallend is dat vrijwel alle studies over fase 3 gaan over docenten die studentreacties analyseren op zwakheden (bv. Abrams & McMillan 2014; Chroinin and Cosgrave_2013; Penuel et al. 2007; Riggan et al 2011; Trauth-Nare & Buck, 2011). Het hebben van observatieschema's of rubrics direct gekoppeld aan de leerdoelen en succescriteria helpt docenten om deze zwakheden te ontdekken en te duiden (Chroinin & Cosgrave 2013). Er zijn slechts een paar studies die expliciet benoemen dat docenten studentreacties ook analyseren op sterke punten of talenten (Brookhart, Moss, & Long, 2010; Conway, 2011; Falk et al., 2007; Matuk et al., 2015). Brookhart et al. ondervinden dat juist het expliciteren en vieren van sterke punten een groot motivatie effect heeft (zie Appendix 2) en Conway laat een mooi concreet voorbeeld zien van hoe het expliciet maken van wat studenten al wel of nog niet weten helpt om gezamenlijk het kennisniveau van de gehele klas op een hoger plan te tillen en om studenten meer autonomie en verantwoordelijkheid in het leerproces te geven (zie Appendix 2). Matuk en collega's beschrijven hoe een docent in een online leeromgeving analyseert hoe ver studenten zijn gevorderd met hun opdrachten. De docent merkt dat een paar studenten bijna klaar zijn, en hij organiseert een apart onderzoeksproject voor deze studenten waarin extra verdieping wordt gevraagd.

Het vijfde thema in deze fase heeft betrekking op de hoeveelheid informatie die de docent verzamelt alvorens tot een conclusie te komen die een vervolgactie in gang kan zetten (naar FT fase 4 of 5). Waar de minder effectieve docent sneller en op basis van minder informatie tot een conclusie komt, zet de effectievere docent meer stappen. Dit kan bijvoorbeeld in de vorm van aanvullend bewijs zoeken, student reacties vergelijken en categoriseren op klasniveau of naar aanleiding van eerste reacties met studenten in gesprek gaan voor verdere aanvulling. Een succesvol bevonden strategie voor het effectiever analyseren en interpreteren van studentreacties was "categoriseren en vergelijken" van studentreacties, zonder daarbij meteen te vervallen in het geven van een oordeel of evaluatie bij ieder studentproduct (Ayala et al., 2008; Falk, 2012; Furtak et al., 2008; Gearhart et al. 2006; Strowbridge 2008; Ruiz-Primo & Furtak, 2007 (zie Box 6); van der Pol et al., 2013). Echter, uit de studie van Ruiz-Primo en Furtak (2007) blijkt dat slechts weinig docenten deze strategie daadwerkelijk toepassen (Ruiz-Primo & Furtak, 2007: zelfs de meest effectieve docent gebruikte deze strategie slechts in 10% van de geobserveerde lessen)). Parr & Limbrick (2008) beschrijven hoe docenten dit in hun klas (basisschool, schrijfonderwijs), samen met hun studenten, doen door gebruik te maken van de fysieke ruimte in de klas. De docenten hangen concrete studentreacties op in de klas en laten de studenten deze vervolgens bekijken en bespreken in het licht van de vooraf gestelde gezamenlijke leerdoelen. Gezamenlijk destilleren ze hieruit aspecten die de klas al wel in de vingers had en onderdelen waarop nog geoefend moest worden.

Hierop werden voor de volgende les de leerdoelen van de klas aangepast en mogen studenten tevens hun individuele leerdoelen aanpassen (Note bene, deze laatste stap valt in fase 5 van de formatieve toetscyclus). Naast individueel analyseren van studentreacties beschrijven sommige studies ook hoe dit in teamverband gebeurt. Goertz en collega's (2009) beschrijven hoe de schoolleider initieert dat alle docenten hun resultaten (uit een volgsysteem) analyseren en zwakke studenten identificeren. De verschillen tussen de klassen worden vervolgens in teamverband besproken alvorens hier consequenties of vervolgacties aan verbonden worden. Een ander voorbeeld wordt gegeven door Buck en collega's (2010) in de context van de lerarenopleiding. De docenten van deze cursus analyseren en evalueren alle beschikbare data continu (in de vorm van actieonderzoek), zoals observaties in de klas, interviews en studentresultaten, en in wekelijkse bijeenkomsten bespreken ze in het docententeam hoe ze de cursus gaan vervolgen en kunnen verbeteren.

Verschillende studies laten zien dat technologie vooral ondersteunend kan zijn bij deze fase en dit soort processen van het vergelijken, categoriseren, aggregeren en interpreteren van student reacties en klasresultaten (Gearhart et al., 2006; Matuk, Linmn, & Eylon, 2015; Penuel et al., 2007; Shirley & Irving, 2015; Simpson Beck, 2011). Student respons systemen (denk aan programma's als Kahoot of Socrative) kunnen individuele student reacties ontlokken en aggregeren op klasniveau. Effectieve docenten gebruiken deze verzameling van studentreacties vervolgens om met de klas te bediscussiëren en analyseren en om op klasniveau makkelijker te ontdekken wat niet goed gaat (Matuk, Linmn, & Eylon, 2015; Penuel et al., 2007; Shirley & Irving, 2015; Simpson Beck, 2011). In de meeste van deze studies gaat het met name over het inzetten van technologie om snel en correct een beeld te krijgen van het zwamheden in de klas om hier direct in de les een vervolg aan te geven. Technologie kan zo het korte, spontane formatieve toetsproces goed ondersteunen. Bij deze studies moet wel genoemd worden dat het hier vooral gaat om makkelijk te meten studentreacties, meestal in de vorm van meerkeuze antwoorden die studenten kunnen kiezen. Enkele studies, meer afkomstig uit de data-based decision making stroming, laten docentpraktijken zien die technologie inzetten voor het uitdraaien van digitale rapporten en het maken van matrices van toetsresultaten, vaak van landelijke summatieve toesten, waarin de scores van de verschillende studenten worden vergeleken (Gearhart et al., 2006; Graney & Shinn, 2005; Goertz et al. (2009), Rice & Smith, 2011).

Bovenstaande voorbeelden laten zien dat studies die over fase 3 gaan, veelal ook over fase 2 gaan. De effectieve docent vervalt niet op basis van de eerste studentreacties in een oordeel, maar vraagt door of zoekt aanvullende studentreacties. De meest succesvolle docenten doen dit, net als eerder beschreven in fase 2, door studenten te activeren zelf in een klasdiscussie hun kennis te expliciteren en uit te diepen, waarbij de docent aanvullende prikkelende vragen stelt die gericht zijn op meer informatie over betreffende leerdoelen te verkrijgen. Hiermee doorloopt een effectievere docent dus een aantal keer de fasen 2 en 3 kortcyclisch direct achter elkaar (Barnhart & Van Es, 2015; Bailey et al., 2014; Conway, 2011; Ruiz-Primo & Furtak, 2007; Shirley & Irving, 2015; van de Pol et al., 2013. Zie box 8), alvorens tot een beslissing te komen voor vervolg (fase 5). Of zoals Ruiz-Primo & Furtak (2006) de praktijk van de meest effectieve docent (Carol) beschrijven:

Box 8. De effectieve docent doorloopt meerdere kleine cycli van fase 2 en 3 in interactiemomenten in de les.

However, these incomplete cycles usually reflected scenarios in which Carol elicited information from multiple students or elicited multiple responses from the same student before acting on the accumulated information. In this way, Carol was able, in a sense, to take the pulse of several students before using the information she had collected, or to help to advance the learning of a particular student

Ruiz-Primo & Furtak, 2006, p.223

Another distinction in Walter's PACT response is how he built on students' science thinking during instruction. For example, he wrote about how he pressed the class during their examination of their pooled data displayed on the board. He questioned students about why different groups' data may be different from each other. The students responded that "individuals are different." As in the previous example, he elaborated on his follow-up response to his students' answer to document how he monitored student progress toward the learning goal: I then furthered the students thinking by asking, "what would be different about people?" The students then came up with responses about different people take in different levels of oxygen, some people are more fit, and different muscle mass. When students would suggest these differences, I

would ask the students to expand on their thinking and why it would make a difference. Therefore, the students were building on their knowledge by connecting the content from lecture to the concepts involved in the lab.

Barnhart & Van Es, 2015, p.89

Typerende docentgedragingen in fase 4: Communiceren met studenten over resultaten.

Opvallend is het geringe aantal artikelen in deze overzichtsstudie dat gericht is op feedback (zie tabel 3), terwijl beide eerdere PPO-NRO overzichtsstudies (Schildkamp et al, 2014; Sluijsmans et al. 2013) feedback als de kern van formatief toetsen benoemen. Er zijn natuurlijk veel studies bekend over kenmerken van goede feedback (e.g. Hattie & Timperley, 2007) en over studentpercepties van goede feedback. Omdat deze studies echter in veel gevallen niet ingingen op concrete formatieve toetspraktijken in de klas en / of op wat de docent concreet DOET, voldeden deze studies niet aan de inclusiecriteria van deze overzichtsstudie. De studies die wel in deze overzichtsstudie zijn meegenomen schijnen daardoor een ander licht op wat de effectieve docent doet in deze fase 4 van de formatieve toetscyclus. Thema's die geïdentificeerd zijn voor deze fase 4 zijn:

1. Doelgerichte en beschrijvende feedback en adresseren van misconcepties
2. Concrete informatie en ruimte voor verbetering
3. Bieden van structuur aan zelf-assessments en peer-assessments

Een effectieve docent in fase 4 geeft doelgerichte en beschrijvende feedback (Chroinin & Cosgrave 2013; Lysaght & O'Leary, 2013) in tegenstelling tot meer oordelende en summatieve feedback (Davis, 2011). De effectieve docent kijkt studentprestaties door de lens van de leerdoelen en geeft ook specifieke feedback die gekoppeld is aan deze leerdoelen. Deze trend komt vaak terug in de kwalitatieve analyse en is mogelijk een verklaring van de relatief hoge co-occurrence van fase 1 en fase 4 in de bestudeerde artikelen (zie tabel 2). Ook in de analyse van fase 4 komt geregeld terug dat docenten beter in staat zijn om specifieke feedback te geven als ze meer vakinhoudelijke kennis hebben (zie ook onderzoeksvraag 3) (Furtak & Heredia, 2014; Herppich, Wittwer, Nückels, & Renkl, 2014; James & Moreland, 2005; Torrance & Pryor, 2001). Tevens laten verschillende studies zien dat wanneer docenten een beter beeld van de leerdoelen hebben en zicht hebben op hoe kennis in dit vakgebied ontwikkelt (i.e. learning progressions), ze ook beter en gericht feedback kunnen geven (Furtak & Heredia, 2014; Herppich et al., 2014; Jones en Moreland, 2005; Lee, 2011). De docenten in de studie van Herppich en collega's laten zien dat zij goed kunnen reageren op onbegrip van studenten als de docent dit zelf opgewekt hebben, maar dat ze het moeilijk vinden om te reageren op een onverwachts, door de student ingebracht kennishiaat. Ook Ayala et al (2008) laat zien dat docenten het moeilijk vinden om on-the-spot te reageren op naïeve student reacties in de klas, wanneer docenten het kennisgebied en de misconcepties die studenten hierin hebben, niet voldoende in de vingers hebben. Dit soort momenten, waarin studenten zelf komen met vragen die een kennishiaat of misvatting laten zien, zijn echter hele krachtige leermomenten ("teachable moments" Ayala et al., 2008), die docenten vaak niet optimaal kunnen benutten. Hulpmiddelen als observatieschema's of rubrics gekoppeld aan leerdoelen en learning progressions kunnen docenten helpen om meer doelgerichte en specifiekere feedback te geven (Furtak & Heredia, 2014; Lee, 2011). Box 9 laat hier een mooi voorbeeld van zien van een docent (basisonderwijs, gym).

Box 9. Voorbeeld van doelgerichte, specifieke en beschrijvende feedback met behulp van observatielijst

One of the teachers described how initially giving feedback consumed her full attention: It was even funny yesterday, how we were out on the green and one of the other older teachers was there with his class and I'd say four times in the space of about thirty seconds 'great ball', 'great ball', 'great ball'. And I was kind of going: all I had in my head was 'great ball' but there's so much more going on here than just 'great ball!' ... (T3, interview 2)

As the lessons progressed, each teacher's feedback gradually became more specific and targeted certain aspects of the children's learning more explicitly: ... structuring the observation using a checklist really focused my feedback to the children and improved the quality of feedback. I was able to give them very specific areas where they were doing well and areas that they needed to improve on. (T5, reflection 6)

These comments highlight the effectiveness of the assessment strategies used and the teachers developing skill to use assessment criteria to diagnose difficulties and support learning

Chroinin and Cosgrave, 2013, p. 228

Een aanvullend cruciaal verschil tussen een meer of minder effectieve formatieve toetspraktijk in de klas is of de feedback concreet gekoppeld is aan ideeën voor verbeteracties (bv. Brookhart et al 2010; Poskitt, 2015; Wylie & Lyon, 2015) en de docent vervolgens ruimte en mogelijkheid geeft om daadwerkelijk iets met deze feedback te doen om verbetering te laten zien (Fluckiger et al., 2010; Hawe & Dixon 2014; Pham & Renshaw, 2015; Restrepo, 2014; Wingate 2010). Een mooi voorbeeld hiervan is te vinden in de “formative assessment cycle” beschreven in de casusstudie van Restrepo (2014, zie Appendix 2). Een kritische aanvulling wordt gegeven door Torrance (2007) die waarschuwt dat feedback en de voorgestelde verbeteracties niet té specifiek moeten zijn, omdat studenten zich anders teveel laten leiden door (gedetailleerde) leerdoelen en feedback, en uit het oog verliezen waar het werkelijk om gaat en/of hun leerproces helemaal laten sturen door de feedback van de docent. Dit is zeker het geval wanneer het doel van de formatieve toetsing gericht is op het stimuleren van een meer zelf-regulerend, autonoom leerproces (Conway, 2011; Davis, 2011). Dit benadrukt het belang van het aanpassen van de feedback aan de doelen die je nastreeft (fase 1 en fase 4)

Het inzetten van zelf- en peer-assessments in deze fase is vaak benoemd als krachtige formatieve toetspraktijk (Sluijsmans et al., 2013). Het cruciale verschil tussen de meer en minder effectieve docent hierin lijkt echter te zijn of de docent een heldere structuur en sturing geeft aan deze zelf- (Restrepo, 2014) en/of peer-feedback (Gerard et al., 2010; Lee, 2011; Paz Demen, 2006; Willis, 2011). Wylie en Lyon (2015) laten duidelijk zien dat het simpelweg bij elkaar plaatsen van peers met de instructie “ga elkaar maar bevragen” of “geef elkaar feedback” niet resulteert in een robuuste implementatie van formatief toetsen. Een experimentele studie van Panadero, Tapia, and Huertas (2012) laat ditzelfde zien bij het gebruik van rubrics. Het simpelweg uitdelen van rubrics leidt niet tot feedback door en voor studenten. Hiervoor zijn aanvullende structuren, formats of sturende vragen nodig. Gerard en collega's (2010) beschrijven bijvoorbeeld hoe docenten een rubric inzetten in een “peer critique” activiteit om studenten te helpen betere peer feedback te geven. Andere gevonden strategieën zijn het geven van prompts (via vragen of voorgestructureerde formats) die linken naar de leerdoelen en criteria (Hawe & Dixon, 2014; Poskitt, 2014; Restrepo, 2014). Een voorbeeld hiervan is de C3B4me (see three (peers) before (you ask) me) strategie (Willis, 2011), waarbij studenten eerst met concrete vragen naar peers moeten gaan alvorens ze naar de docent komen. Een student in de studie van Restrepo (2014), waarin studenten in verschillende cycli zichzelf beoordelen (zie ook Appendix 2) en deze zelf-beoordeling vervolgens met elkaar en de docent bespreken in een “conference on performance” werkvorm, rapporteert over het belang van gestructureerde zelf-assessment het volgende:

T: Do you think making self-assessments is effective?

S: I think it is, I know what I'm failing in. I think that without doing self-assessment as proposed in the classroom (i.e. eerst zelfstandig en dan in conferentie vorm), one is not able to self-asses on his own.

T: How do you feel talking to me when we compared your self-assessment with the assessment I completed about your performance in one task or during the whole course?

S: Usually, we agreed on what I had spotted like my weaknesses. It was like confirming what I had to work on.

T: Did you think that talk was effective? S: Yes it was. Because if one lies, one does not value what one really has to work on and one would go on regardless that. One has to confront oneself and with the teacher one analyzes, one has to encourage oneself to overcome those difficulties and that is an incentive to learn (Restrepo, 2014,. P. 176)

Typerende docentgedragingen in fase 5: Vervolgacties ondernemen

In fase 5 waren drie thema's te onderscheiden in de kwalitatieve analyse:

1. Docenten hebben moeite met het daadwerkelijk aanpassen van instructie gebaseerd op de eerdere fasen;
2. Ondernomen vervolgstrategieën gericht op het aanpassen van het onderwijs / de instructie, en
3. Ondernomen vervolgacties om bepaald (leer)gedrag van de student te stimuleren.

Ten eerste bleek dat veel docenten moeite hebben met fase 5. Deze fase kreeg in de kwantitatieve analyses, gevolgd door fase 2, de meeste NOT codes (zie tabel 3) en uit de kwalitatieve analyse bleek duidelijk dat dit de moeilijkste stap is van de formatieve toetscyclus, een stap waarin docenten nog vrij weinig handelingsrepertoire hebben (Forbes et al., 2015; Gearhart & Osmundson, 2008; Goertz et al., 2009; Gottheiner & Siegel, 2010; Herman et al., 2015; Sabel, Forbes, & Zangori, 2015; Wylie & Lyon, 2015) en/of een stap die in weinig studies concreet wordt gemaakt. Vaak blijft het bij terminologie als “naar aanleiding van de geïdentificeerde zwakheden, past de docent vervolgonderwijs aan” (e.g. “*Carol acted on information accumulated over multiple ESRU cycles....*” Ruiz-Primo & Furtak, 2006, p. 223). Suurman en collega’s (2010) beschrijven bijvoorbeeld voorbeelden van docenten (Claire, Terry en Angela) die experimenteren met leuke, doelgerichte methodieken om studentreacties te ontlokken (fase 2), hoe zij deze analyseren (fase 3), maar vervolgens staat voor fase 5 alleen omschreven *dat* deze docenten informatie uit deze nieuwe methodieken halen die hen helpt keuzen te maken voor vervolgonderwijs, maar niet *hoe* ze dat doen en *wat* ze dan veranderen (zie Box 10). De studie van Ayala en collega’s (2008) laat ook zien dat de timing soms een probleem is: de docenten in deze studie zagen de (informeel bedoelde) formatieve momenten toch al toetsen met een summatief karakter die ze mee naar huis namen om na te kijken. Daardoor misten ze de “teachable moments”, namelijk om direct in te springen op de geïdentificeerde misconcepties en hun instructie aan te passen.

Verschillende van deze studies laten ook zien dat docenten zelf aangeven fase 5 een moeilijke stap te vinden, terwijl ze zich wel realiseren dit het een belangrijke stap is, en graag iets willen doen met de studentreacties die ze hebben opgemerkt:

I’ve made changes in instruction based on student evidence but I am not sure they are the best changes—I at least try” en “now that I know where they are, what do I do? [how do I] make changes in instruction based on assessment results?” (docent in Gearhart & Osmundson, 2008, p. 31)

Box 10. Voorbeelden van geen concrete uitwerking van fase 5: aanpassing van het onderwijs

Teachers’ use of assessment also provides them with information about how to proceed with learning activities. Claire commented about her quizzes providing her with a sense of what students can do and what steps she might take next. In a similar way, Terry described how she used student responses to adjust her pedagogical actions when students were confused about her use of colours in an activity and Angela commented on how student responses in math journals help her to adjust her teaching.

Suurman et al. (2010, p 414)

Wylie & Lyon (2015) observeerden docenten in de klas en ‘scoorden’ hun formatieve toetspraktijken met de rubric waarmee gekeken werd of docenten zinvolle formatieve methodieken inzette om informatie van studenten te ontlokken die formatief gebruikt konden worden (FT-fase 2), of ze deze informatie ook analyseerden (FT-fase 3) en of ze de analyse gebruikte om hun instructie aan te passen (FT-fase 5). In 46% van de gevallen doorliepen docenten zowel fase 2, 3 als 5 na een twee jarig professionaliseringstraject (zie ook onderzoeksvraag 4). Hierbij bleek bovendien dat de kwaliteit van de instructie aanpassing niet altijd hoog was. Barnhart en Van Es (2015) gebruiken een vergelijkbare rubric (zie Figuur 3) voor de mate waarin docenten aandacht hebben voor studentbegrip (attending, fase 2), ideeën van studenten analyseren (analyzing, fase 3), en hierop reageren (responding, fase 5). Ook deze studie laat zien dat docenten over het algemeen lager scoren op “responding” dan op attending en analyzing, ook na een professionaliseringstraject. Ook de zeer recente studies van Herman en collega’s (2015) en Sabel et al (2015) concludeerden ook dat docenten, ook na interventies, “het identificeren van goede vervolgstappen” (FT- Fase 5) het minste uitvoerden of van lage kwaliteit en gebaseerd op verkeerde analyses van studentreacties. Dit geldt ook voor een studie in de lerarenopleiding, waar leraren-in-opleiding weinig strategieën implementeerden op hun stageplek, ook al was hun begrip van de FT-cyclus wel toegenomen na een cursus gericht op de FT-cyclus (Buck et al., 2010).

Figuur 4. Rubric voor het scoren van “noticing skills” van docenten (Barnhart & Van Es, 2015)

Table 2
Levels of sophistication for noticing skills.

Skill	Low sophistication	Medium sophistication	High sophistication
Attending	Highlights classroom events, teacher pedagogy, student behavior, and/or classroom climate. No attention to student thinking.	Highlights student thinking with respect to the collection of data from a scientific inquiry (science procedural focus).	Highlights student thinking with respect to the collection, analysis, and interpretation of data from a scientific inquiry (science conceptual focus).
Analyzing	Little or no sense-making of highlighted events; mostly descriptions. No elaboration or analysis of interactions and classroom events; little or no use of evidence to support claims.	Begins to make sense of highlighted events. Some use of evidence to support claims.	Consistently makes sense of highlighted events. Consistent use of evidence to support claims.
Responding	Does not identify or describe acting on specific student ideas as topics of discussion; offers disconnected or vague ideas of what to do differently next time.	Identifies and describes acting on a specific student idea during the lesson; offers ideas about what to do differently next time.	Identifies and describes acting on a specific student idea during the lesson and offers specific ideas of what to do differently next time in response to evidence; makes logical connections between teaching and learning.

Als tweede thema is een variatie aan aanpassingsstrategieën gevonden die docenten toepassen in hun klas en de mate waarin ze deze gekozen vervolgstategie baseren op een eerdere analyse van studentreacties. Een aantal studies laat zien hoe docenten hun instructie doelgericht aanpassen naar aanleiding van geïdentificeerde kennishiaten of misconcepties in de gehele klas (Falk et al., 2012; Furtak & Heredia, 2014; Gerard et al., 2010; Matuk et al., 2015; Mitton-Kukner et al., 2014; Phelan et al., 2012). Falk en collega's beschrijven hoe docenten in het basisonderwijs hun kennis over instructiestrategieën (zie ook onderzoeksvraag 3) gebruiken om vervolgstategieën te kiezen, maar ook om te evalueren hoe eerdere instructiestrategieën de reacties van studenten op formatieve momenten hebben beïnvloed, en hoe instructie en formatieve assessments bij elkaar aansluiten. Mitton-Kukner en collega's (2014) beschrijven een mooi citaat van een docent die aangeeft haar instructie op verschillende manieren aan te passen, al naar gelang wat ze opmerkt in de discussies van studenten (zie Box 11) en Gerard en collega's (2010) beschrijven hoe groepen docenten in een professionaliseringstraject gezamenlijk hun lesplan aanpassen op basis van hun analyse van studentwerk. Zo merken de docenten bijvoorbeeld dat hun studenten “plate boundary” (aardrijkskunde) goed kunnen uitleggen, maar dit vervolgens weer vergeten. Daarom ontwerpen ze een werkblad waarop studenten een schets met uitleg moeten maken, en dit moeten bewaren en gebruiken bij het gehele verdere project.

Box 11. Twee voorbeeld van een docent die instructie aanpast op basis van eerdere analyse

I found [literacy strategies] gave me the ability to gauge where the students are, especially with the entrance and exit slips and the reflections, cause then, based on what they say or if they are able to answer a particular question, I can see if they're confused, or if they're getting it, or if they choose to ask me a question. . . . And if it's just one student in particular then I can talk to them independently, but if I see a pattern then I can address the problem as a class...If I did find that they were having difficulty, I may not be able to proceed with the next lesson as I had planned, I might have to backtrack and address certain issues. So, I found on a day-to-day basis I could be influenced . . . depending on what they had said. Just from circulating and seeing, observing what they were doing, kind of gave me an idea of what I need to spend more time on. (Pre-service science and physical education teacher Mary, May 1, 2013).

Mitton-Kukner et al. (2013), p. 403

She also noticed that “students explain each plate boundary clearly, and usually can identify a landform associated with that type of boundary, but don't remember this critical information later! (Activity 3, Step 4)” (Lilly workshop artifact, Yr 2). Lilly decided that, “students need to be guided to look at the models of plate boundaries and convection currents more carefully” (Lilly reflection, Yr 2). Lilly and colleagues created a lesson plan that distributed student use of Plate Tectonics over 22 class periods as shown in Table 6. The lesson plan incorporated additional lab activities (Days 8–9, 15, 16), supplemental worksheets (Days 13 and 16), and homework assignments (Days 9 and 16). Each new activity guided students to revisit Plate Tectonics visualizations multiple times.

Gerard et al. (2010)

De meest voorkomende geïdentificeerde aanpassingsstrategieën waren gericht op klasniveau. Hierbij ging het voornamelijk om (1) het tempo aanpassen (pacing) of (2) de stof of de les herhalen (reteaching). Pacing slaat op het afwijken van een bepaald vastgelegd tempo in het curriculum door bijvoorbeeld sneller of langzamer door bepaalde stof heen te gaan (Riggan & Nabors Oláh, 2011; Shirley et al., 2014). Brookhart en collega's (2010) laten zien dat het een hele stap was voor docenten om af te durven wijken van het voorgestructureerde curriculum en zich ook bekwaam te voelen om dit te doen. Pas na een langdurig professionaliseringstraject (Zie appendix 3) ontdekten docenten dat zij hun lessen veel krachtiger konden maken door soms af te wijken van het voorgeschreven tempo en onderwerpen als de analyse van studentreacties hier aanleiding toe gaven, en ontwikkelden zij ook een gevoel van bekwaamheid (i.e., self-efficacy) om dit te doen. Reteaching was veelal een oppervlakkige strategie, van het op dezelfde manier opnieuw vertellen van de inhoud. Docentpraktijken verschilden in of gekozen vervolgactiviteit gebaseerd was op eerder geïdentificeerde zwakheden op klasniveau, of dat de aanpassingen eigenlijk weinig samenhang met eerdere analyse liet zien (Abram & McMillan, 2014; Fluckiger et al, 2010; Gerard et al., 2010; Gottheiner & Siegel, 2010; Matuk et al., 2015; Shirley et al., 2014; Trauth-Nare & Buck, 2011; Wall, 2012). Een aantal studies liet echter zien dat docenten lessen herhaalden zonder dit te baseren op eerdere analyses van studentreacties (Ateh, 2015; Graney 2005; Herman et al 2015; Penuel et al 2007; Sabel et al., 2015; Schneider & Andrade, 2013). Tevens laten verschillende studies zien dat docenten hun vervolgonderwijs niet aanpassen aan eerder geïdentificeerde misconcepties (Buck, TrauthNare, & Kaftan, 2010; Forbes et al., 2015; Goertz et al., 2009; Gottheiner & Siegel, 2010), en dit ook niet koppelen aan te behalen leerdoelen (Abram & McMillan, 2014; Sabel et al., 2015). Forbes en collega's (2015) beschrijven bijvoorbeeld hoe docenten wel studentreacties ontlokken en kennishiaten identificeren, maar hun aanpassingen in de instructie baseren op strategieën die ze al kennen, en niet aansluitend op de geïdentificeerde hiaten. Ook Goertz en collega's (2009) beschrijven hoe maar een klein deel van de docenten in hun studie bij hun re-teach strategieën probeert om de inhoud ook op een andere manier over te brengen. De meeste docenten herhalen de inhoud, en doen dit ook door middel van dezelfde instructie strategieën.

Shirley & Irving (2015) en Matuk en collega's (2015) laten daarentegen zien dat technologische systemen deze stap, het identificeren van een effectieve vervolgstategie op basis van eerdere analyses, kunnen ondersteunen. Matuk en collega's (2015) beschrijven bijvoorbeeld hoe een elektronische leeromgeving de docent in staat stelt om (on)begrip van groepen studenten te monitoren, en daarop instructie aan te passen, bijvoorbeeld door in de volgende les de uitleg aan het begin van de les aan te passen of de elektronische leeromgeving even stil te leggen voor een extra uitleg (zie ook Box 13).

Box 12. Voorbeeld van hoe technologie fase 3 en 5 kan faciliteren

In another example, Mrs. Roberts used CCT to uncover a confusing aspect of position-time graphs. She used CCT to gather responses to a question prompt and found that many students had difficulty interpreting a position-time graph for an object that did not start its motion immediately.

One thing that I noticed was that, from what I saw with the Quick Polls is that if the reference point or the starting point isn't right when the time starts, [the students] don't understand how the graph is going to change. So that's something we will have to review tomorrow. We can use the CBR, and whether they do it in groups, or we do it as a demonstration, then they need to see that just because the time starts, the object doesn't have to start moving. And I saw, from the Quick Poll, that many of them did not get that. (Mrs. Roberts: POI)

Mrs. Roberts's comments indicated that she was beginning to formulate a revised lesson plan for the next day's lesson, incorporating specific activities that aim to clarify this particular point of confusion.

Shirley & Irving, 2015, p 64.

Slechts een klein aantal studies liet zien dat docenten gingen differentiëren naar individuele (Abrams & McMillan, 2014; Matuk et al., 2015; Van der Pol et al., 2013) of kleine groepen studenten (Gearhart & Osmundson, 2008; Goertz et al., 2009; Matuk et al., 2015; Phelan et al., 2012; Wall, 2012). De studie van Abrams en McMillan (2014) liet zien dat individuele aanpassingsstrategieën meestal buiten schooltijd vielen. Met andere woorden, docenten gaven deze studenten extra huiswerk of vroegen extra voorbereidingsactiviteiten die de student buiten de les moesten doen. In de studie van Goertz en collega's (2009) geven de docenten een lessenserie van ongeveer 5 weken, en gebruiken ze de 6^e week om herhalingsactiviteiten en verdiepingsactiviteiten aan te bieden voor verschillende studenten, die daarvoor steeds op een verschillende manier werden gegroepeerd. In docenthandelen was weinig terug te zien hoe docenten hun instructie aanpasten voor hoogvliegers: *"differentiating at the high end does not receive much attention"* (Wall, 2012). Brookhart en collega's (2010) benoemen hoe hun docenten "de sterke punten van hun studenten expliciteren en vieren" en wat dit voor motiverend effect op studenten heeft (zie ook Appendix 3). Dit wordt echter niet opgevolgd door een andere aanpak of ander leertraject. Conway (2011) beschrijft daarentegen een mooi voorbeeld van hoe een docent dit wel kan doen en in de lessen de talenten van iedere student goed kan benutten. De studie van Matuk en collega's (2015) laat zien hoe docenten in de elektronische leeromgeving kunnen zien hoe bepaalde studenten erg voorlopen in de opdrachten. Voor deze studenten ontwikkelde de docent een extra verdiepende onderzoeksopdracht.

Het derde thema van fase 5 gaat niet over het aanpassen van het onderwijs, maar over stimuleren van ander leergedrag van de student. Deze code ("Phase 5, stimulate student to change their learning behaviour" of "stimulate students to regulate their own learning") is maar zeer mondjesmaat terug gevonden (8 respectievelijk 26 keer gescoord). Wel is een aantal studies gevonden die beschrijven hoe de docent de studenten actief aan het werk zet door eerst studentreacties uit te lokken (fase 2), die eventueel te analyseren samen met de studenten, en de studenten vervolgens in groepjes aan het werk zet om verschillende antwoorden, oplossingen of verklaringen te bespreken (Kang et al., 2014) en hun eigen werk aan te passen (Forbes et al., 2015; Kang et al., 2014; Yin et al., 2014). De focus ligt hier steeds op het verbeteren van eigen eindproduct en niet op stimuleren van bepaald leergedrag (zelfsturing, zelf-regulatie). Twee geïdentificeerde kritische factoren voor docenten die met formatieve toetspraktijken zelfgestuurd/zelf-regulerend leren stimuleren lijken (1) het al dan niet bieden van structuur voor hoe de student zou moeten veranderen of het eigen leren zou moeten reguleren en (2) de gerichtheid van de docent om een autonome, zelfverantwoordelijke student te willen stimuleren (Conway, 2011; Davis, 2011; Restrepo, 2014; Wylie & Lyon, 2015). Davis (2011) laat zien in een studie die expliciet gericht is op het stimuleren van zelf-regulerend leren, docenten in hun lessen geen aandacht besteden aan het aanzetten van studenten tot zelf-regulerend gedrag. De docenten blijven gericht op makkelijk meetbare, feitelijke kennis. Hun leeractiviteiten, en vervolgactiviteiten na toetsresultaten, zijn daarmee ook niet gericht op het stimuleren van zelf-regulerend gedrag. Enkele studies, vooral studies waarin concrete docentcasussen beschreven worden, geven voorbeelden van docenten die wel activiteiten inzetten om studenten te stimuleren hun leergedrag te veranderen en hier verantwoordelijkheid voor te nemen (e.g. Conway, 2011; Kearney, 2013; Restrepo, 2014; Suurman et al., 2010). Deze studies hebben gemeen dat de betreffende docenten het ontwikkelen van een autonome student, die verantwoordelijkheid neemt voor het eigen leren, hoog in het vaandel hebben staan. Zij zijn voortdurend bezig met het prompten van (peer)studenten om naar hun eigen leren en prestaties te kijken, door middel van meer of minder creatieve methodieken en formats, om hierbij passend leergedrag te laten zien (zie box 13).

Box 13. Voorbeelden waarmee docenten studenten aanzetten tot het veranderen van eigen leergedrag.

I think I'm ready for level 11 because

I think I'm not ready for level 11 because

For overcoming my difficulties I need to:

Restrepo (2010)

Colour groups

On occasion, Angela used colour groups in her classroom as a way to differentiate her instruction based on the students' self-assessment of their needs. She established four student groups, each with a separate colour, signifying the level of assistance needed, ranging between requiring no, little, guided, or significant teacher assistance. She did not assign students to the groups but asked students to determine how they were faring

with a particular concept and to place themselves in the group that best reflected their immediate need with the given set of problems:

The skills that I'm hoping to develop when I do this are self-advocacy, because they have to self-assess first. Let's say, 'Here's wherever I am and here's where I think I want to be and what I am going to need', and then self-advocacy 'Can I ask for that help?' (Angela, interview)

The group of students that indicated they were confident and comfortable with the work could also assist members of another group who were less confident yet willing to try the problems. Students could also change groups if they thought they had incorrectly assessed their needs. This technique not only helped to promote self-assessment and agency within the students but also allowed Angela to focus her attention on those students who required additional assistance.

Suurman et al., 2010, p. 406-407

De gehele Formatieve Toetscyclus en de kleine 2-3-5 cyclus

Uit de beschrijvingen van de losse fasen hierboven, kwam al meerdere keren naar voren dat de fasen niet geheel los van elkaar te bekijken zijn. Vanuit het theoretische kader is bovendien ook wenselijk dat docenten gedragingen uit de gehele FT-cyclus laten zien. In geen enkele studie geselecteerd voor deze overzichtsstudie

kwamen alle vijf de fasen van de FT-cyclus terug en in geen enkele studie lieten docenten dus gedrag zien passend bij de gehele FT-cyclus. Wel werd er in meerdere studies expliciet vanuit een cyclische gedachte naar formatieve toetspraktijken in de klas gekeken (bv. Buck, Traught-Nare, & Kaftan, 2010; Forbes, Sabel, Biggers, 2015; Ruiz-Primo & Furtak, 2006, 2007; Sabel, Forbes, & Zangori, 2014; van der Pol, Volman, & Beishuizen, 2012). Zo keken Ruiz-Primo & Furtak (2006; 2007) bijvoorbeeld naar docentpraktijken vanuit de ESRU-cyclus (Elicit-Student Response-Recognize-Using), Van der Pol en collega's vanuit de cyclus van contingent teaching en Forbes et al. en Sabel et al. vanuit een model gericht op student-responsieve instructie. Ondanks dat al deze cycli net iets anders zijn, hebben ze met elkaar gemeen dat ze allemaal gaan over de fasen 2, 3 en 5 van onze Formatieve Toets-cyclus: opwekken van informatie, interpreteren van student reacties en het plannen, de conclusies gebruiken om een nieuwe strategie voor vervolginstructie te plannen of in te zetten. De studies verschilden in of ze meer formeel en geplande (Buck et al.; Forbes et al.; Sabel et al.) of meer informele, ongeplande methodieken gebruikten, zoals "assessment conversations" (Buck et al., Ruiz-Primo & Furtak; van der Pol et al., 2013; 2014). Deze laatste set studies beschreef allemaal meer en minder succesvolle docentcasussen, waarbij de meer succesvolle docent gekenmerkt werd door:

1. Meer volledige cycli (dus meer 2-3-5, ipv 2-3 of 2-5 of alleen 2)
2. Meer onvolledige cycli van response opwekken (fase 2), eerste interpretatie (fase 3), doorvragen (fase 2), aanvullende reactie krijgen, verschillende studentreacties samenklusteren (fase 3)....., bij voldoende informatie komen tot een gegronde vervolgstap (fase 5) (dus 2-3-2-3----> 5)
3. Vragen stellen en doorvragen op conceptueel begrip in plaats van het goede antwoord
4. Meer student sturing toelaten in (klas)discussies, waarbij de docent inspringt op wat studenten zeggen en inbrengen in plaats van de docent die stuurt

De studies van Ruiz-Primo & Furtak toonden bovendien ook aan dat de docenten die dit het beste deden, het meeste effect hebben op de prestaties van hun studenten. Studenten scoorden significant hoger op kennis en 'performance' toetsen (redeneervragen).

De studies van Forbes en Sabel zijn professionaliseringsstudies (zie verder Onderzoeksvraag 4), die laten zien dat docenten wel groeien in het goed kunnen ontlocken van student begrip (fase 2) en beter worden in het evalueren van student antwoorden (fase 3), maar het kiezen van een geschikte vervolgstategie gebaseerd op hun gevonden analyses bleef erg lastig (fase 5). Buck en collega's ontdekten vooral hoe moeilijk student-docenten, ondanks gericht onderwijs in formatief toetsen, het bleven vinden hun formatieve praktijken gericht te laten zijn op het conceptuele begrip van studenten en meer sturing van de student te laten komen in een meer gelijkwaardige relatie en discussie.

Samenvattend: Deze studies (1) leveren dus bewijs voor de effectiviteit van de kleine cyclus 2-3-5 (zie box 6, 8 en 9); (2) benadrukken twee cruciale factoren van de effectievere docent. Ten eerste, gericht zijn op (conceptueel) begrip in plaats van het goede antwoord en, ten tweede, meer student-sturing oproepen en inspringen op wat studenten inbrengen, en (3) laten zien hoe moeilijk fase 5 blijft (het ondernemen van vervolgacties) ook na professionalisering.

Naast deze studies die expliciet vanuit een cyclisch raamwerk naar formatieve toetspraktijken van docenten keken, zijn er nog meer studies gevonden die in de resultaten docentgedrag in de 2-3-5 cyclus laten zien (bv. Conway, 2011; Gerard, Spitulnik, & Linn, 2010; James & Moreland, 2005; Mitton-Kukner & Orr, 2014, Penuel et al., 2007; Suurman et al., 2010). In tegenstelling tot de meeste van bovenstaande expliciet cyclische studies, maakten deze studies ook veel gebruik van meer formele, of in ieder geval meer geplande, assessment methodieken. Voorbeelden hiervan zijn mind-maps (Mitton-Kukner & Orr, 2014), complexe probleem-oplostaken (Suurman et al., 2010) of inzetten van digitaal student response systeem (Penuel et al., 2007). In Box 14 een voorbeeld van een docent uit de studie van Gerard et al (2010) die de fasen 2-3-5 doorloopt met behulp van een meer formele, geplande formatieve toetsmethodiek (het plaat tektoniek project).

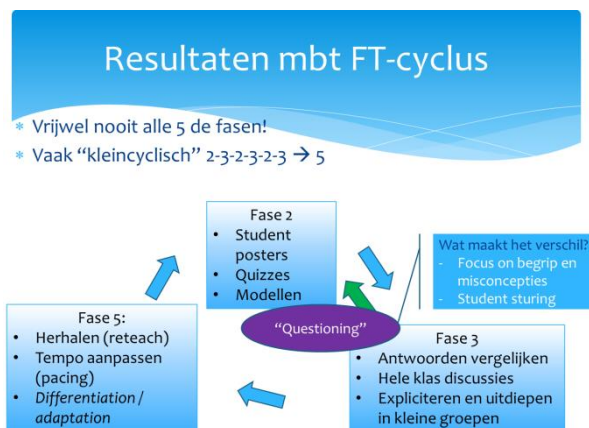
Box 14. Voorbeeld van een docent die de kleine FT-cyclus (2-3-5) doorloopt met een formeel, geplande formatieve toets

Ted made only a few, but quite significant customizations to the Plate Tectonics project. Ted compared his Year2 Plate Tectonics embedded assessment data to his Year 1 data. He found that, "students did better than last year from beginning to end [of the project]. But even though some finished early this year they still do not have clearly defined and detailed answers. . . indicating they were not fully grounded in the plate tectonics concepts" (Ted reflection, Yr 2). Ted linked students' performance to a "need to require more explanations in student answers, more asking why" (Ted reflection, Yr 2). He subsequently added two activities to the end of the Plate Tectonics project. One activity, for example, presented images of a dramatic local geological event, with embedded assessments asking students to use their knowledge of plate tectonics to explain why this geological event was happening in their backyard and what it meant to their future. Ted hoped this activity would make the project more relevant to his class and in turn, motivate students to investigate the why underlying geological events.

Gerard et al., (2010), p.

Deze studies geven vaak aan dat effectieve docenten het belang inzien van de gehele FA-cyclus, en hun attitude ten opzichte van formatief toetsen daardoor laten leiden. Ook in deze studies vertonen docenten die als meest effectief geïdentificeerd worden gedrag dat zich kenmerkt door de kleine cyclus (2-3-5) of de 2-3-2-3-2 --> 5 cyclus. Dus deze bevinding lijkt sterk bevestigd in verschillende onderwijscontexten, vakgebieden en met voor zowel informele, spontane als formele, geplande formatieve toetsing. Figuur 5 geeft een schematische weergave van deze kleine cyclus.

Figuur 5. De kleine 2-3(-2-3)-5 cyclus voor effectie(ver) formatief toetsen



Onderzoeksvraag 2: Studenteffecten en relaties met docentgedragingen

Soorten effecten

Onderzoeksvraag 2 was de volgende: "Welke effecten op studenten hebben de toetspraktijken van docenten in termen van leeractiviteiten, leerprocessen en/of leeruitkomsten?" Voor de beantwoording van deze vraag is ten eerste geteld in hoeveel artikelen een effect op studenten werd gemeten, onafhankelijk van de gehanteerde onderzoeksmethode en onafhankelijk van het soort effect (leeractiviteit / leerproces / leeruitkomst). Dit bleek het geval in 42 van de 106 artikelen in deze reviewstudie: in bijna 40% procent van de artikelen was sprake van het meten van een effect op studenten.

De gemeten effecten zijn vervolgens gecategoriseerd in 5 verschillende soorten effecten:

- Effecten op kennis(ontwikkeling) van studenten
- Effecten op (verbeteren van) schrijven door studenten
- Effecten op motivatie en/of attitudes van studenten
- Effecten op leerprocessen zoals zelf-regulatie en het zoeken van advies/feedback

- Effecten op (algemene) studentpercepties en/of sfeer in de klas

In de meeste artikelen was sprake van het meten van effecten op de kennis(ontwikkeling) van studenten. Dit werd op verschillende manieren gemeten door middel van kennistoetsen, vaak door landelijke toetsen of state-wide tests in de VS met mc-vragen en open vragen (Beesley, 2009; Brookhart et al., 2010; Gerard et al., 2010; Wall, 2012; Wiliam, 2004). Een groot aantal van deze studies heeft betrekking op science onderwijs, en daar werd vaak het conceptueel begrip van studenten gemeten (Aschbacher, 2006; Ruiz-Primo et al., 2006) en vaak werd naast een toets van begrip van science-concepten ook een experimentje gevraagd, waarbij studenten de uitkomst moesten voorspellen of interpreteren (bijvoorbeeld bij drijven / zinken, elektrische circuits; Furtak et al., 2008; Ruiz-Primo et al., 2007; Yin et al., 2014). In een deel van de studies werd ook gemeten of studenten beter in staat waren om wetenschappelijke redeneringen te geven voor geobserveerde fenomenen (b.v. Heller et al., 2012; Yin et al., 2014), zoals in de studie van Kang en collega's (2014) waarin "good scientific explanations" in het schriftelijk werk van studenten werd geanalyseerd. In één studie werd conceptueel begrip van technologie gemeten op een andere manier dan door middel van kennistoetsen: in de studie van Jones en Moreland (2005) werd dit gemeten door middel van het observeren van conversaties van studenten en het analyseren van het werk van studenten. In studies naar taal / schrijven werd meestal gekeken naar de kwaliteit van het schrijfwerk van studenten, soms door middel van een pre-test en post-test (Paz Demeo, 2006; Poskitt, 2014). In de studie van Wingate (2010) werd onderzocht hoe studenten hun teksten verbeteren door te analyseren hoe studenten aan de slag gingen met de feedback van de docent (56% van de studenten verbeterden hun cijfer door de feedback).

Een heel andere categorie zijn de studies waarin effecten werden gemeten over studentmotivatie, betrokkenheid en/of attitudes. De studies van Beesley (2009), Lee (2011) en Patel et al., (2012) lieten positieve effecten zien op studentmotivatie, gemeten door middel van een vragenlijst. In de studie van Lee (2011) werden studenten ook geïnterviewd over hun attitude ten opzichte van schrijven. In de interviews voorafgaand aan de studie gebruiken studenten woorden zoals "angst" en "nooit meer", terwijl studenten aan het einde van het jaar waarin de docent formatief toetsen implementeerde uitspraken deden zoals "ik ga mijn best doen" en "180 graden verandering". Brookhart en collega's (2010) beschrijven docentervaringen met formatief toetsen, waarbij de docent rapporteert dat de studenten meer gemotiveerd waren, meer gefocust op de taak en meer betrokken omdat de FT-praktijken een gevoel van controle gaven over het eigen leerproces. Ook Kay en Knaac (2009) rapporteren positieve effecten op betrokkenheid en motivatie als reactie op een audience response system, waarin studenten via stemkastjes of mobiele telefoons multiple choice vragen beantwoorden waarvan de antwoorden direct op het scherm in de klas geprojecteerd werden. De docent ging naar aanleiding hiervan direct met studenten in discussie over wat de gevonden resultaten lieten zien.

De vierde categorie effecten op studenten heeft betrekking op effecten op zelf-regulatie, zelf-monitoring en het zoeken naar advies of feedback op het eigen werk. In één studie (Panadero et al., 2012) werd dit gemeten door middel van vragenlijsten; in de andere studies door middel van observaties en/of interviews met studenten. Alle studies rapporteerden positieve effecten op studenten. Uit de studies van Jones & Moreland (2005), Poskitt (2014) en Restrepo (2014) bleek dat studenten meer in staat waren om hun eigen knowledge gaps te identificeren, omdat ze zich meer bewust waren van de beoogde leerdoelen en deze leerdoelen beter begrepen. Het gaat hier om de capaciteit om het eigen kennen en kunnen in te schatten (zelf-evaluatie) en op basis hiervan het eigen leren te kunnen sturen. Ook waren studenten meer bereid om advies te vragen. Ook uit de studie van Lorente en Kirk (2013) bleek dat studenten zich meer bewust waren van hun eigen fouten en betere manieren konden vinden om zich te verbeteren, doordat ze meer verantwoordelijkheid krijgen over het beoordelingsproces en meer eigen beslissingen konden nemen. In deze studie werden studenten expliciet betrokken bij het formuleren van leerdoelen en het kiezen van toetsmethoden om dit aan te tonen. Ook de studie van Pham en Renshaw (2015) laat interessante resultaten zien: studenten waren beter in staat om vragen te stellen en advies te vragen aan hun peers en de docent en de vragen die ze stelden werden ook steeds beter van kwaliteit in termen van complexiteit en diepgang. Door de FT-praktijken van de docent zagen

ze leren meer als een collectief proces waarin gezamenlijk kennis wordt geconstrueerd en vragen stellen dus belangrijk is voor leren (in plaats van dat kennis alleen uit het boek afkomstig is). Tot slot heeft de studie van Willis (2011) een focus op het inzetten van formatief toetsen om student autonomie te vergroten door formatief toetsen te zien als een participatief en sociaal-cultureel proces waarin de student een “sense of belonging” en identiteit als lerende ontwikkelt.

De laatste categorie van gerapporteerde effecten op studenten heeft betrekking op algemene studentpercepties over bijvoorbeeld het belang of nut van formatief toetsen (Lee, 2011; Newby & Winterbottom, 2011) of positieve ervaringen (Lorente & Kirk, 2013) en op effecten op klas/schoolniveau, zoals een positievere sfeer in de klas (Fluckiger et al., 2010; Pham & Renshaw, 2014). Fluckiger en collega's beschrijven bijvoorbeeld hoe FT-praktijken leiden tot beter klasdiscussies en een betere leercultuur met meer vragen en opmerkingen van studenten. Ook Pham en Renshaw beschrijven hoe de klascultuur verandert: studenten zijn meer betrokken bij elkaars werk, delen meer kennis en geven constructieve feedback op elkaar, een grote verandering ten opzichte van de gebruikelijke praktijk in Vietnam waarbij klasdiscussies worden geleid en gedomineerd door de docent.

Relatie tussen docentgedrag en effecten op studenten

Naast een overzicht van de gemeten effecten op studenten, is voor de beantwoording van onderzoeksvraag 2 gekeken naar de code “teacher behaviour => effect on student” omdat in deze fragmenten uit de studies zichtbaar wordt welk concreet docentgedrag invloed had op studenten (in de vorm van welk effect dan ook). In appendix 2 is een overzicht te vinden van deze studies, met de concrete docentgedragingen, bijbehorende fasen uit de FT-cyclus, en de effecten op studenten die werden gerapporteerd. Deze paragraaf geeft een samenvatting van de belangrijkste bevindingen, geordend volgens de soort effecten op studenten (zie hierboven) en de docentgedragingen (eventueel per fase) die daarbij gevonden werden. Het belangrijkste verschil werd hierbij gevonden tussen studies gericht op kennis(ontwikkeling) van studenten en studies gericht op processen zoals zelf-regulatie en autonomie.

In de studies waarbij een effect werd beoogd op de kennis(ontwikkeling) van studenten, hebben verreweg de meeste docentgedragingen betrekking op kleine FT-cyclus, of op de cyclus van fasen 2-3-2-3, eventueel gevolgd door fase 5. Typische voorbeelden van deze studies – waarbij positieve effecten werden gemeten op de kennis van studenten – zijn de studies in het science onderwijs (Aschacher 2006; Heller et al., 2014; Ruiz-Primo, 2006, 2007; Yin et al., 2014). Docentgedragingen in deze studies zijn te kenmerken door het ontlokken van studentreacties, vaak door middel van een experimentje (Predict-Observe-Explain), gevolgd door vergelijking en discussie tussen studentantwoorden in kleine groepjes, terugkoppeling van de groepjes naar de gehele klas, gevolgd door klasdiscussies (Yin et al., 2014). Ook het toepassen van de ESRU-cyclus (elicit student response – student response – recognize student response – use information to adapt teaching) valt onder de docentgedragingen in deze categorie. Het onderzoek van Herman en collega's (2015) laat meer in het algemeen zien dat docenten effecten bereiken bij studenten wanneer zij méér gebruik maken van FT-strategieën: lesdoelen koppelen aan curriculumdoelen, meer gevarieerde methodieken gebruiken om studentreacties te ontlokken, meer tijd nemen om studentwerk te analyseren, feedback geven aan individuele studenten en de hele klas, en meer inhoud herhalen (re-teach). Pham en Renshaw (2015) geven een wat andere invulling aan de 2-3-2-3-5 FT-cyclus, maar rapporteren ook positieve effecten op het begrip van studenten. Een docent (casus) zet een summatieve mc-test formatief in door studenten elkaars antwoorden in groepjes te laten vergelijken en verschillen aan elkaar te laten uitleggen. Vervolgens kiest de groep het beste alternatief en lichten alle groepjes hun keuze klassikaal toe. Ook in de studies van Mitton-Kukner (2014) en Gerard en collega's (2010) wordt een variant op de 2-3-2-3-5 FT-cyclus ingezet, en worden positieve effecten gerapporteerd op de kennis(ontwikkeling) van studenten. Een interessante toevoeging wordt gegeven door Hickey en Anderson (2007) in hun onderscheid tussen verschillende (tijds)niveaus in FT-praktijken van docenten. Uit deze studie blijken positieve effecten op de kennis(ontwikkeling) van studenten wanneer docenten kunnen schakelen tussen deze niveaus, bijvoorbeeld close-level (tijdsspanne van een dag) met

informele toetsen als leidraad om activiteiten vorm te geven, en discourse-level (tijdsspanne van minuten binnen de les) om discussies gericht te houden op de beoogde leerdoelen.

Naast deze studies die betrekking hadden op fasen 2-3-5 werd nog een aantal studies gevonden waarin (bijna) geen aandacht was voor fase 3 (analyseren), maar waarin wel positieve effecten werden gerapporteerd op de kennis(ontwikkeling) van studenten (Herrpich et al., 2014; Kang et al., 2014; Phelan et al., 2012). De studie van Herrpich en collega's (2014) gaat over tutor-gesprekken (fase 2-4), waaruit blijkt dat studenten hoger scoren op een mc-toets wanneer hun tutor meer feedback geeft op basis van geconstateerde fouten, getriggerd door de tutor, of spontaan door de student. Uit deze studie blijkt vooral hoe belangrijk het is dat tutores fouten herkennen en hierop reageren (wat moeilijk blijkt bij fouten die de tutor niet zelf heeft ontlokt, maar spontaan door studenten worden ingebracht). De studie van Kang en collega's (2014) heeft weer een andere focus, namelijk op het gebruik van verschillende scaffolds in FT-taken, zodat de student in zijn/haar reactie gestuurd wordt richting het beoogde doel (fase 1-2). Ook hier blijkt dat de kwaliteit van de "explanations in science" toenam als de docent meer en betere scaffolds inzet. Concluderend blijkt uit deze studies een aantal concrete docentstrategieën in met name fasen 2-3-5 en wat concrete strategieën in fasen 1 en 2 waarmee duidelijke effecten worden bereikt op de kennis(ontwikkeling) van studenten.

Naast studies die effecten beoogden op de kennis(ontwikkeling) van studenten, was een aantal studies gericht op het bevorderen van zelfsturend leren, autonomie en/of het kunnen inschatten van het eigen kennen en kunnen. Deze studies laten (bijna) allemaal zien dat de docent daartoe bewuste FT-strategieën inzet, meestal gericht op het doelgericht en gestructureerd betrekken van studenten in één of meer fasen van de toetscyclus (Conway, 2011; Willis, 2011; Pham & Renshaw, 2015; Brookhart et al., 2010; Fluckiger et al., 2010; Kearney, 2013; Lee, 2011; Lorente & Kirk, 2013). Studies waarin positieve effecten worden gerapporteerd, laten een heel ander patroon aan docentstrategieën zien dan de studies gericht op kennis(ontwikkeling). Een duidelijk voorbeeld is het onderzoek van Lorente en Kirk (2013) waarin de docent eerdere ervaringen met toetsing bespreekt met studenten, en vervolgens studenten leerdoelen en toetsvormen laat bedenken die deze problemen kunnen voorkomen. Criteria en toetsmethodieken worden bewust samen met studenten ontwikkeld. Hetzelfde is terug te zien in de onderzoeken van Kearney (2013) en Conway (2011) waarin studenten een actieve rol wordt gegeven in het formuleren van eigen leerdoelen met als doel autonomie, het plannen van het eigen leertraject en "sustainable learning" te bevorderen. In de studies van Willis (2011) en Torrance and Pryor (2001) wordt hieraan toegevoegd dat het belangrijk is om ook leerdoelen te formuleren gericht op participatie (Willis, 2011) of sociale norm (Torrance & Pryor, 2001), en ook in dit onderzoek worden studenten op een gestructureerde manier uitgedaagd om actief deel te nemen. Al deze studies laten positieve effecten zien op de ervaringen van autonomie en het actief sturen van het eigen leren.

Studies waarin wordt gestreefd naar zelfsturend leren laten ook zien dat het betrekken van studenten het beste werkt op een gestructureerde manier, bijvoorbeeld door actieve participatie te bevorderen door het (verplicht) stellen van vragen aan elkaar (Phan & Renshaw, 2015), en de methode C34 (see three peer students before asking me; Willis, 2011). Fluckiger en collega's (2010) beschrijven de three color quiz, waarbij studenten actiever worden betrokken door en gestructureerde activiteit waarbij de student eerst met zwart schrijft wat hij zelf weet, de groep dit aanvult met groen, en blauwe teksten worden aangevuld met informatie uit tekstboeken. Ook op deze manier wordt de participatie van studenten zichtbaar gemaakt, met als resultaat betere participatie en een beter klasklimaat. Tot slot wordt in deze studies veel gebruik gemaakt van peer-assessment en zelf-assessment als methoden om studenten te betrekken bij de FT-cyclus, liefst op een gestructureerde manier door bijvoorbeeld feedbackformulieren met focuspunten (Kearney, 2013; Lee, 2011; Wylie & Lyon, 2015).

Onderzoeksvraag 3: Docentcompetenties voor formatieve toetsgedragingen

Onderzoeksvraag 3 was: *"Welke competenties hebben docenten nodig om goede formatieve toetspraktijken in klas neer te kunnen zetten?"* Zoals al meerdere malen benadrukt, kijkt deze studie alleen naar de vraag over

wat docenten moeten kennen en kunnen vanuit de concrete (effectieve) formatieve toetspraktijken die ze laten zien. Hiervoor werd de code “docentcompetenties” 162 keer gescoord, verdeeld over 56 artikelen, oftewel 53% van de studies. Dit zijn studies die vanuit beschrijving van de concrete toetspraktijk komen tot cruciale onderliggende formatieve toetscompetenties (46%) of studies die een expliciete professionaliseringsinterventie uitvoeren gericht op een aantal docentcompetenties (54%). Deze studies geven veel informatie over welke toetscompetenties cruciaal zijn, wel goed aangeleerd kunnen worden en waar docenten juist, ook na professionalisering, moeite mee blijven houden. In deze resultaten sectie worden de geïdentificeerde docentcompetenties benoemd en waar mogelijk gekoppeld aan een of meerdere fasen van de FT-cyclus en gedragingen die in de studies gevonden zijn die hierdoor beïnvloed worden. NB. Deze gedragingen zijn in de meeste gevallen eerder beschreven bij onderzoeksvraag 1. Voor de leesbaarheid, worden in deze paragraaf achter de gevonden activiteiten waarin bepaalde competenties zijn teruggevonden, geen referenties genoemd. Deze zijn altijd terug te vinden in onderzoeksvraag 1.

Doorleefd begrip. Ten eerste is het cruciaal dat de docent een *doorleefd begrip van formatief toetsen* heeft ontwikkeld. Dit wil zeggen dat de docent de FT-cyclus daadwerkelijk begrijpt en doorleefd en woorden kan geven aan hoe formatief toetsen in de klas eruit ziet (Moss et al., 2013), in plaats van formatief toetsen ziet als het inzetten van een instrument/strategie/methodiek waarvan studenten potentieel zouden moeten kunnen leren (e.g. Rice & Smith, 2011; Robinson, Myran, Strauss, & Reed, 2014; van der Pol, Volman, & Beishuizen, 2013). Robinson et al laten zien dat het hebben van een doorleefd conceptueel begrip van formatief toetsen, waarbij docenten ook inzien hoe hun FT gedrag invloed heeft op studentgedrag en leren, voorwaardelijk is voor het gebruik van verschillende FT activiteiten uit de FT-cyclus (i.e., *communiceren* van doelen, studenten betrekken, verschillende methodieken inzetten, data analyseren). Het verkrijgen van dit inzicht vraagt een “belief change”. Een veranderde opvatting over toetsen waarbij docenten toetsen, en het verzamelen en analyseren van student reacties, moeten gaan zien als een krachtig middel voor het verbeteren van zowel het leerproces als het onderwijsproces (Rice & Smith, 2011).

Vakinhoudelijke kennis en learning progressions. Twee veelvuldig genoemde competenties, nauw aan elkaar verbonden, is het *kunnen toepassen van vakinhoudelijke kennis ten behoeve van formatief toetsen* (Aschbacher & Alonzo, 2006; Falk, 2012; Forbes et al., 2015; Herman et al., 2015; Sabel et al., 2015; Talanquer et al., 2013; 2015), en het *hebben van inzicht in hoe studenten hun kennis in het vakgebied ontwikkelen en wat veel voorkomende misconcepties zijn* (oftewel: inzicht in “learning progressions” Bv. Falk, 2012; Furtak & Heredia, 2012; Gearhart et al., 2006). Vooral in studies die uitgevoerd zijn in het “science” domein, kwam dit veelvuldig terug, waarbij verschillende studies aantonen dat inhoudelijke vak kennis voorspellend was voor formatieve toetspraktijken van docenten (Herman et al., 2015; Feldman & Capobianco, 2008; Robinson, 2014; Sabel et al (2015). Uit een veelheid van studies kwamen vakinhoudelijke kennis en inzicht in learning progressions naar voren als cruciale voorwaarde voor het vormgeven van alle fasen van formatief toetsen. Hieronder staat een opsomming van hoe vakinhoudelijke kennis en inzicht in learning progressions in de kwalitatieve analyse terug gevonden zijn in de concrete docentactiviteiten in de verschillende FT-fasen: Docenten met meer vakinhoudelijke kennis en zicht op de learning progressions, kunnen:

- op een heldere manier leerdoelen en succescriteria formuleren (fase 1)
- assessment taken ontwerpen of vragen stellen die daadwerkelijk conceptueel begrip van het vakgebied ontlokken, in plaats van gericht zijn op het vinden van het goede antwoord (fase 2; ref)
- inspringen op vragen, reacties of discussies van studenten in de les, die een misvatting of kennislak laten zien (fase 2-3)
- studentantwoorden analyseren interpreteren op daadwerkelijke (conceptueel) begrip in plaats van op oppervlakte kenmerken (zoals vorm en presentatie) (fase 3)
- uit studentantwoorden foutief begrip en misconcepties identificeren (fase 3)
- doelgerichte en specifieke feedback die terugkoppeld op de doelen en vervolg ontwikkelstappen (learning progressions) die studenten doorlopen (fase 4)

- structuur bieden voor zelf- en peer-feedback gekoppeld aan doelen en het expliciteren en uitdiepen van begrip (bv. “just right” guidance van Aschbacher & Alonzo, 2006)

Vakdidactisch handelingsrepertoire. Naast vakinhoudelijke kennis, wordt het kunnen inzetten van vakdidactische kennis (pedagogical content knowledge) bij het vormgeven van formatieve praktijken ook veel aangehaald (Forbes et al., 2015; Herman et al., 2015; Sabel et al., 2015). Docenten moeten groeien in hun vakdidactische repertoire en hier zelf bewustzijn over ontwikkelen (bv. wat doe ik en waarom? Torrance & Pryor, 2001). Deze competentie relevant voor alle fase van de FT-cyclus, maar in het bijzonder voor fase 2, het kunnen inzetten van gepaste assessment methodieken, en fase 5, het kunnen kiezen en inzetten van een effectieve vervolgstراتيجية, en voor het creëren van alignment (aansluiting) tussen de verschillende fasen. Docenten met een goed vakdidactisch handelingsrepertoire, kunnen:

- de FT-toets cyclus als doelgerichte en samenhangende cyclus uitvoeren (gehele cyclus; Tang et al., 2010). Dat wil zeggen, dat docenten in staat zijn om alignment tussen de verschillende toetsfasen te creëren. Specifiek is dit gevonden voor:
 - Assessment methoden inzetten die passen bij de doelen en learning progressions (fase 1-2)
 - Student antwoorden analyseren op doelen en kernconcepten van het vakgebied (fase 1-2-3)
 - Feedback koppelen aan doelen en learning progressions (fase 1-4)
 - Vervolgstrategieën kiezen gebaseerd op gevonden analyses en misconcepties (fase 3-5) (bv.
- een rijker repertoire van assessment methodieken gebruiken en deze flexibel toepassen (fase 2). Specifiek is dit gevonden voor:
 - flexibel en spontaan kunnen inspringen op vragen, reacties of discussies van studenten in de les, die een misvatting of kennisgat laten zien (fase 2-3) Dit is vooral cruciaal voor het vormgeven van informeel formatief toetsen door middel van bv. assessment conversations Herppich et al., 2014; Offerdahl & Tomanek, 2011; Ruiz-Primo, 2011).
 - goede vragen (open, conceptueel en gekoppeld aan leerdoelen) kunnen stellen en goede (klas/groeps)discussies te kunnen voeren, leiden en stimuleren
 - structuur geven aan peer-/zelf assessments die studenten zelf en onderling uitlokken om hun begrip te expliciteren en uit te diepen gekoppeld aan leerdoelen.
- Gepaste vervolgstراتيجيةën identificeren en toepassen voor zowel zwakkere als sterkere student en geïdentificeerde misconcepties (fase 5)

Een aanvullend punt bij deze competentie, dat in een aantal studies een verklarende factor in de resultaten blijkt te zijn, is of docenten *durven afwijken van het vooraf vastgelegde curriculum* (Aschbacher & Alonzo, 2006; Brookhart et al., 2010; Forbes et al., 2015; Furtak & Heredia, 2014) en of *docenten zich bekwaam voelen om af te wijken van dit voorafgeplande curriculum en methode* (Brookhart et al., 2010).

Student-docent relatie met student-sturing. Een vierde geïdentificeerde competentie die in veel studies het verschil leek te maken tussen goede en nog betere formatieve toetspraktijken is: *het tot stand kunnen brengen van een gelijkwaardige docent-student relatie waarbij de studenten sturend zijn* aan het formatieve toetsproces, zoals de discussies en activiteiten in de klas waar de docent vervolgens flexibel op kan inspringen (bv. Feldman & Capobianco, 2008; Lysaght & O’Leary, 2013; Parr & Limbrick, 2010; Pham & Renshaw, 2015; Ruiz-Primo & Furtak, 2006; 2007; TrauthNare & Buck 2011; Wylie & Lyon, 2015).

It was also observed that the more encouraging verbal behaviours adopted by the teachers created a more relaxed studying atmosphere, and closed the gap between the teachers and the students. The students joked and laughed more often with the teachers and with each other. Lessons toward the end of the semester were transformed into informal discussions, in which the group members and the teachers engaged in sharing relevant knowledge and providing each other with constructive feedback. This was far removed from formal content-based talks managed by the teachers in the first lesson (Pham & Renshaw, 2015)

"We found that FA is not simply a cognitive process for eliciting students' understanding, rather it is a fundamentally relational act in which students and teacher can participate, construct, and coordinate classroom processes. (TrauthNare & Buck, 2011, p. 395)".

Buck en collega's (2010) laten zien dat student-docenten met die "hoogste niveau" van formatief toetsen laten zien, inzicht demonsteren in wat dit voor relationele student-docent processen vergt:

"Formative assessment can only be successful when students are comfortable enough to let what they don't know show and that will only happen if they trust their teacher. Additionally, through working with the students in a trusting, positive relationship, students are more willing to ask questions and express concerns, and in so doing they are able to take ownership of their learning and the teacher can craft lessons which appeal to the students and are responsive to their needs." (Emily, CS, 5/02)

Om dit te bereiken moet de docent een leeromgeving weten te creëren waarbij studenten en docenten samen participeren, kennis construeren en samenwerken. Docenten die meer effectieve formatieve toetspraktijken lieten zien, konden:

- een veilige omgeving creëren, van openheid, gelijkwaardigheid, betrokkenheid en een "sense of democracy" de norm was (Feldman & Capobianco, 2008; Pham & Renshaw, 2015; Restrepo, 2014; van der Pol, Volman, & Beishuizen, 2012)
- expliciet met studenten in gesprek gaan over student-docent relaties en rollen (bv. in gesprek gaan over "Testing vs. Helping questions", Torrance & Pryor, 2001; TrauthNare & Buck, 2011)
- studenten ondersteunen in het leren goede vragen te stellen (Bloxham & Campbell, 2010)
- de fysieke ruimte inzetten om actieve studentparticipatie te weeg te brengen (bv. *Tafels in groepen ipv in lange rijen* (Rice & Smith, 2011); *Antwoorden van studenten in de klas ophangen en er met z'n alle bij gaan staan om te bespreken en te vergelijken* (Parr & Limbrick, 2007) Rice & Smith, 2011; Parr & Limbrick, 2007)

Wanneer docenten formatief toetsen willen inzetten om autonomie, verantwoordelijkheid en zelf-sturend vermogen bij studenten te stimuleren, in tegenstelling tot formatief toetsen inzetten als bijvoorbeeld voorbereiding op de summatieve toets of om minder summatief te gaan toetsen, dan is deze student-sturing nog belangrijker. Willis (2011) verwoordt dit mooi in het volgende citaat:

Understanding how to attend to, evaluate and enhance student learning through their conversations and participation, and how students negotiate meaning and an identity from their participation in AfL practices, is a challenging new form of assessment literacy for teachers that underpins the spirit of AfL (p. 412)

Docenten die dit konden bereiken,

- stelden leerdoelen op gericht op participatie, samenwerking en/of sociale norm (fase 1)
- boden studenten ruimte (binnen bepaalde kaders) om hun eigen leerdoelen te formuleren en mede assessments te ontwikkelen / leeractiviteiten te bepalen (fase 1)
- zetten gestructureerd peer / self-assessments in, waarmee studenten hierin oefenen (fase 5)

Inzetten van technologie voor ondersteuning van het formatieve toetsproces. Technologie wordt in een deel van de artikelen genoemd als mogelijkheid om formatief toetsen te ondersteunen (b.v. Kay & Knaack, 2009; Matuk et al., 2015). Hierbij gaat het meestal over een elektronisch scoringssysteem, waarbij studenten via stemkastjes of telefoon antwoorden op multiple choice vragen kunnen invoeren, waarvan daarna direct op het scherm de klasscores zichtbaar worden. De studie van Matuk en collega's laat een uitgebreidere elektronische leeromgeving zien die effectief wordt ingezet voor het ondersteunen van het formatief

toetsproces in de klas. En in een klein aantal artikelen gevonden in deze overzichtsstudie moeten docenten digitale rapportages over studenten en klassen, veelal van landelijke toetsen, kunnen interpreteren en gebruiken in vervolgonderwijs. *Docenten moeten die technologie kunnen hanteren ten behoeve van formatief toetsen in de klas* (Feldman & Capobianco, 2008; Fluckiger et al., 2010; Matuk et al., 2015; Penuel et al. 2007; Shirley & Irving, 2015). Docenten die dit goed konden, konden

- student reacties clusteren en vergelijken en individuele / klas zwakheden en sterke punten identificeren (fase 3)
- studentantwoorden inzetten als basis voor klas/groepsgesprekken (fase 2/3),
- individuele en groepsfeedback te geven, veelal in de digitale omgeving (fase 4)
- snel (meestal in dezelfde les) de gevonden zwakheden/misconcepties adresseren in een gepaste vervolgactiviteit (fase 5)

Studenten uit de studie van Kay & Knaack (2009) rapporteren allerlei positieve leer -en motivatie-effecten van het werken met het Audit Response System in de klas voor formatieve doeleinden, terwijl ze het gebruik ervan voor summatieve doelen niet goed vinden. Dat benadrukt dus de toegevoegde waarde van een dergelijk systeem sterk leunt op of de docent het goed formatief kan inzetten.

Onderzoeksvraag 4: Professionaliseringstrajecten en effecten

Onderzoeksvraag 4 was: *“Aan welke formatieve toetscompetenties (ozv3) en gedragingen (ozv1) besteden professionaliseringstrajecten aandacht? Op welke gebieden zijn deze succesvol en waarop (nog) niet?”* In 43 van de 106 studies in deze overzichtsstudie werd expliciet een professionaliseringsinterventie op formatief toetsen ingezet met als doel de formatieve toetspraktijken van docenten in de klas te versterken. Appendix 3 geeft een uitgebreid overzicht van deze professionaliseringstrajecten, hun invulling, de FT-fasen waar ze expliciet op gericht waren, de FT competenties die expliciet werden aangesproken en de effecten die zichtbaar waren in de praktijken van de docenten. Tevens is in deze tabel (in Appendix 3) een aantal instrumenten expliciet gemarkeerd die zeer behulpzaam kunnen zijn bij of mooie voorbeelden bieden voor anderen die formatief toetspraktijken van docenten willen ‘scoren’ of ontwikkelen. Voorbeelden hiervan zijn de rubrics van Poskitt (2014) of Sato, Wei, & Darling-Hammond (2008) die beide concrete docentpraktijken in de klas en student-docent interacties in kaart brengen en ontwikkeling hierin laten zien. Deze paragraaf geeft een samenvatting van de belangrijkste bevindingen van de professionaliseringstrajecten op de vijf fasen van de formatieve toetscyclus die uit de kwalitatieve analyse gekomen zijn.

Een duidelijk zichtbaar patroon uit Appendix 3 was dat een beduidend aantal studies professionaliseert op de kleine FT-cyclus: 2-3-5. Op verschillende manieren gaan docenten in deze verschillende studies aan de slag met het ontwikkelen van geschikte toetsmethoden, het analyseren en interpreteren van studentantwoorden en het plannen, identificeren en soms uitvoeren van een logische vervolgstap. Uit het merendeel van deze studies blijkt dat docenten het makkelijkste groeien in alleen fase 2 of in fase 2 en 3, terwijl fase 5 in vrijwel alle studies niet ontwikkelde (Sabel et al., 2015) en/of niet gebaseerd werd op een eerdere analyse geleerd uit fase 3 (bv. Feldman & Capobianco, 2008). Geen enkele studie professionaliseert expliciet op alle vijf de fasen van de FT-cyclus. Aschbacher en Alonzo (2006) komen hier het meest bij in de buurt. In hun succesvolle studie besteden ze veel aandacht aan het stimuleren van alignment tussen de verschillende fasen (doelen bij methode, analyse bij doelen, vervolgstap bij analyse etc), waarbij fase 1-2-3 de meeste expliciete aandacht krijgen, terwijl fase 5 minder expliciet aan bod komt. Fase 4 komt ook aan bod, maar als een apart blok gericht op “wat is effectieve feedback” losgekoppeld van de FT-cyclus en de andere fasen. Effecten op docenten zijn vooral zichtbaar op fase 1, 2 en 3.

Ten aanzien van Fase 1 zijn de resultaten gemixt. Sommige studies rapporteren dat docenten inderdaad betere leerdoelen en succescriteria gaan formuleren (bv. Furtak & Heredia, 2014; Gearhart et al., 2006; Torrance & Pryor, 2001), andere studies rapporteren geen groei in deze fase (bv. Beesley et al., 2009; Wylie & Lyon, 2015),

of wel toename in kwantiteit maar met lage kwaliteit (ze doen het meer, maar de implementatie is van slechte kwaliteit, bv. Furtak et al., 2008). Het gezamenlijk leren maken van een learning progression lijkt bij te dragen aan het beter werken met leerdoelen en het beter inzetten van vervolgstappen van formatief toetsen (Furtak & Heredia, 2014; Panizzon & Peggs, 2008) en ook professionaliseringstrajecten die expliciet inzetten op het belang van alignment tussen de doelen en de volgende formatieve toetsfasen lijken succesvol (Aschbacher & Alonzo, 2008; Chroinin & Casgrave, 2013; Falk, 2012; Gearhart et al., 2006; Panizzon & Peggs, 2008).

Fase 2 lijkt het meest effectief gestimuleerd te kunnen worden. Verschillende studies rapporteren dat docenten na professionalisering een rijker repertoire van assessment methodieken inzetten of een groter repertoire van vragen-stel technieken kunnen toepassen (bv. Feldman & Capobianco, 2008; Sato, Wei, Darling-Hammond, 2008; Traught-Nare & Buck, 2011). Echter, professionaliseringstrajecten die alleen inzetten op fase 2, ook wel de meer instrumentele benadering van formatief toetsen genoemd (Coffey, Hammer & Levin, 2011; Sluijsmans, Joosten-Ten Brinke, & Van der Vleuten, 2013), leiden niet tot enig effect (Offerdahl & Tomanek, 2011; Mui So & Hoi Lee, 2011; Tang et al., 2010). Met name studies die inzetten op het koppelen van assessmentmethodieken aan bepaalde doelen die worden nagestreefd, lijken effect te hebben op het uitbreiden van het repertoire van de docent (bv. Falk, 2012; Chroinin & Cosgrave, 2008). Barnhart en Van Es (2015) laten vervolgens zien dat deze fase 2, het goed kunnen ontlocken van een student reactie, voorwaardelijk is om de analyse (fase 3) en vervolgstap (fase 5) goed te kunnen doen, terwijl het goed kunnen uitvoeren van fase 2, geen garantie is voor het ook goed kunnen analyseren en opvolgen van deze resultaten.

Enkele studies zijn gevonden waarin de focus ligt op fase 3 (Rice & Smith, 2011; Goertz, Olah, Riggan, 2009; Graney & Smith, 2005; McNaughton, Lei, & Hsiao, 2012). Hieruit blijkt ook dat alleen professionalisering op het leren analyseren en interpreteren van student/klasresultaten weinig effect heeft als dit niet gekoppeld is aan actieve professionalisering op hoe hier vervolgens mee om te gaan in het onderwijs (fase 5). Rice & Smith laten heel kleinschalig, en McNaughton et al. grootschalig, zien dat gezamenlijk nadenken over hoe interpretaties te vertalen zijn naar vervolgonderwijs en differentiatie noodzakelijk zijn om te leiden tot effect zichtbaar in de klassenpraktijk van de docent én vervolgens daadwerkelijk (meetbare) effecten op studentresultaten sorteren. Verder blijkt uit de kwalitatieve analyse dat het analyseren van een variatie van studentantwoorden (bij voorkeur werk van eigen studenten) effectief lijkt om beter te leren analyseren van studentreacties (Falk, 2012; Gearhart et al. 2006; Heller et al., 2012; Herman et al., 2015). Dit hield met name in dat docenten “dieper” gingen kijken naar studentantwoorden dan alleen naar oppervlakte kenmerken en de analyses gingen verbinden aan de doelen die ze nastreefden en de kernconcepten van het vakgebied.

Met betrekking tot fase 4 is ook in deze studies, net als bij de eerdere onderzoeksvragen, niet veel gevonden.

De bevindingen voor fase 5 zijn helder: hier ligt het grootste professionaliseringsvraagstuk. In veel studies uit Appendix 3 is aandacht voor fase 5 en in de grootste meerderheid van deze studies blijven docenten ook na professionalisering handelingsverlegen als het gaat om het kiezen van een gepaste vervolgstategie. De studie van Gearhart & Osmundson (2008) is hierop een uitzondering

Twee aanpalende thema's die uit de kwalitatieve analyse van de professionaliseringstrajecten kwamen waren:

- Het ontwikkelen van een doorleefd begrip van formatief toetsen is vaak een geïntegreerd of inherent onderdeel van de professionalisering. Verschillende van deze studies rapporteren ook een verandering in de opvattingen van de docenten over toetsing (bv. Offerdahl & Tomanek, 2011; Moss et al., 2013; Robinson, Mayran, Strauss, & Reed, 2014; Sato, Wei, & Darling-Hammond, 2008)
- Professionaliseringstrajecten die expliciet inzetten op zowel vakinhoudelijke als vakdidactische kennisontwikkeling lijken succesvol te zijn (bv. Falk, 2012; Heller et al., 2012; Herman et al., 2015; Jones & Moreland 2005), in ieder geval voor fase 2 en 3. Ook verschillende studies die niet expliciet professionaliseren op vakken, rapporteren in de resultaten dat duidelijke gebleken is dat het hebben dat vakinhoudelijke kennis voorspellend (b)lijkt te zijn voor het verbeteren van de formatieve

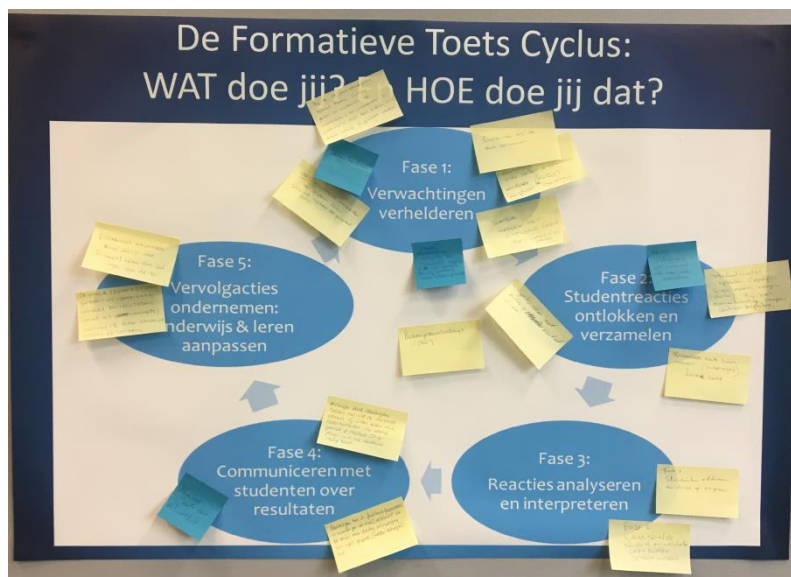
toetspraktijk van de docenten. Deze studies concluderen veelal ook dat concretere aandacht voor vakdidactische kennis en handelingsrepertoire noodzakelijk is voor met name het beter kunnen uitvoeren van fase 5: ondernemen van vervolgstappen passend bij de analyses van studentantwoorden.

Hoofdstuk 5. Kennisdeeldag

Dit hoofdstuk legt verslag van de kennisdeeldag, waarin de resultaten van de overzichtsstudie zijn besproken met het onderwijsveld. De doelen van deze bijeenkomst waren ten eerste om de resultaten te delen, ten tweede om docenten te prikkelen om met deze resultaten kritisch op hun eigen formative toetspraktijken in de klas te reflecteren en ten derde om elkaar te inspireren met concrete voorbeelden bij een of meerdere FT-fasen. Door in verschillende werkvormen met deze doelen aan de slag te gaan, konden we zicht krijgen op de mate waarin docenten de resultaten van deze reviewstudie herkenden, welke resultaten ze konden toepassen en op welke manier, en welke discussiepunten dit opleverde. Het laatste doel van de bijeenkomst was om gegevens te verzamelen die konden helpen bij het formuleren van de praktische implicaties voor verdere professionalisering, passend bij vraagstukken die echt leven bij docenten.

Hieronder worden eerst de uitgevoerde activiteiten beschreven alsmede de verschillende databronnen die dit opleverde. Daarna wordt de dag samengevat in een aantal reacties en discussiepunten die de kern van de reactie van het veld op de resultaten van deze reviewstudie weergeven.

Figuur 6. Voorkennisactivatie: bij de opstart beschrijven docenten (goede) voorbeelden van wat zij doen in één van de FT-fasen



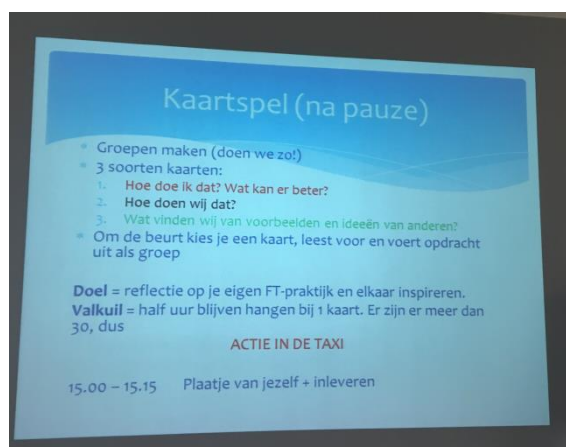
Tijdens een kennisdeeldag (17 februari 2017) zijn 52 deelnemers (van de 64 aanmeldingen) aan de slag gegaan met de resultaten uit deze reviewstudie. De deelnemers waren voornamelijk docenten, maar ook beleidsmakers, onderwijskundigen en adviseurs en vertegenwoordigden PO, VO, (V)MBO, HBO, WO, Ministerie OCW en onderwijsadviesbureaus. Als input voor de dag zijn posters gemaakt van de vijf FT-fasen waarin per fase beschreven staat:

- 1) Wat DOET de effectieve docent in deze fase?
- 2) Welke vragen stel je jezelf en elkaar?

Na een voorkennisactivatie (zie figuur 6) werden in een korte powerpoint presentatie de belangrijkste bevindingen uit de overzichtsstudie gedeeld. Vervolgingen de deelnemers actief aan de slag met de vijf fasen van de FT-cyclus via een kaartspel (zie Figuur 7) waarin zij reflecteerden ze op hun eigen FT-praktijk, elkaar inspireerden en concrete voorbeelden uit de artikelen in de overzichtsstudie (zie ook de boxen in deze rapportage) bespraken. Hierbij konden zij ook gebruik maken van de 5 FT-fasen posters (zie Appendix 4) die per

fase de kern beschrijven van wat de effectieve docent DOET en welke vragen de docent zichzelf en elkaar zouden moeten stellen om samen kritische te reflecteren op eigen formatieve toetspraktijken.

Figuur 7. Instructie voor het kaartspel rondom de FT-cyclus en de 5 fasen



Na afloop van het kaartspel maakten de deelnemers “een plaatje van zichzelf” (zie figuur 8). Ze scoorden zichzelf op de vijf fasen van de FT-cyclus en beschreven per fase wat ze in hun klas al heel goed doen en wat ze willen leren. Tevens beschreven ze een concreet voorbeeld dat inspirerend kan zijn voor andere docenten (als voorbereiding op de bijeenkomst was al gevraagd hierover na te denken). Deze “plaatjes van jezelf” resulteerden in rijke beschrijvingen van wat docenten doen in hun klassen en wat ze nog willen leren. De beschreven voorbeelden worden gebundeld, geormerkt met schooltype/onderwijsniveau, en als praktische output van deze overzichtsstudie overdraagbaar gemaakt.

Figuur 8. Doenten maken “een plaatje van zichzelf” op de 5 FT-fasen

	Fase 1: verhelderen van doelen en succescriteria	Fase 2: ontlokken en verzamelen van data van studenten	Fase 3: analyseren en interpreteren van student reacties/antwoorden	Fase 4: communiceren over resultaten met studenten	Fase 5: Zetten van vervolgstappen - Lessen aanpassen - Anders studentgedrag stimuleren
Welke score zou je jezelf geven? 1 = dit doe ik nog helemaal niet 5 = dit doe ik heel goed in mijn klas	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Beschrijf per fase wat je al goed doet in jouw klassen?					
Beschrijf per fase wat je ontdekt hebt dat je nog zou moeten/willen leren					
Beschrijf een inspirerend voorbeeld van hoe jij in een klas aan formatief toetsen doet Wij verzamelen al jullie voorbeelden in een bestand en					

De laatste activiteit van de middag bestond uit een korte presentatie met discussie en groepsopdracht over de resultaten met betrekking tot “Wat moet de docent kennen en kunnen?” en “Waar is verdere professionalisering nodig?” en aan het einde beschreven de deelnemers hun tip, top en eye-opener.

Al deze activiteiten leverden input over hoe docenten de resultaten van deze studie ervaren, herkenden, nieuw vonden en konden gebruiken. Hieronder een samenvatting van de belangrijkste en voor de studie relevante reacties en discussiepunten.

Belangrijkste reacties en discussiepunten uit het veld

Reacties. De meest genoemde eye-opener refereerde naar de “learning progressions” en de aandacht voor misconcepties hierin. Hier had nog vrijwel niemand eerder van gehoord, terwijl docenten uit het PO zich hier vanuit hun onderwijspraktijk het meeste bij konden voorstellen. Het scheen voor veel mensen nieuw licht over formatief toetsen en men zag verschillende mogelijkheden voor zich over hoe learning progressions behulpzaam kunnen zijn bij het beter vormgeven van meerdere fasen van de FT-cyclus.

Een andere eye-opener was het expliciet benoemen van leerdoelen gericht op participatie of sociale norm - naast inhoudelijke leerdoelen - als de docent met formatief toetsen het doel heeft om autonomie, zelf-sturing en/of zelf-regulering te stimuleren (Torrance & Pryor, 2001; Willis, 2011). Hieruit bleek ook duidelijk dat de deelnemers verschillende doelen hadden om formatief toetsen in te willen zetten, waarbij het stimuleren van zelf-regulatie lang niet altijd aan de orde was.

De volgende onderdelen werden expliciet door docenten benoemd als heel herkenbaar: het belang van het stellen van heldere leerdoelen, alhoewel de meesten dit alleen deden aan het begin van het vak in de studiewijzer; inzetten van eenvoudige technologie (bv. kahoot) om formatieve discussies in de klas aan te wakkeren; de rol die van studenten wordt verwacht in formatief toetsen, maar ook het probleem om studenten in deze actieve rol te krijgen; de focus op zwakheden; het belang van vakkennis om formatief toetsen goed te kunnen vormgeven; de moeilijkheid van het stellen van goede vragen die niet reproductie gericht zijn (dat is: op zoek naar het goede antwoord), het kiezen voor nog eens uitleggen (re-teach) als vervolgstراتيجية, en de handelingsverlegenheid als het gaat om kiezen van verschillende en gedifferentieerde vervolgstراتيجиeën.

Fase 3 (analyseren en interpreteren van student reacties) werd door veel docenten in hun zelfbeoordeling aangegeven als een fase die ze niet doen, niet bewust doen, of niet eerder hadden gezien als onderdeel van formatief toetsen.

Zowel in de discussies als in de zelfbeoordelingen (plaatje van jezelf) bleek een grote behoefte aan kennis, vaardigheden en inspiratie om meer (individuele) feedback en differentiatie vorm te geven in grote klassen en binnen beperkte tijd.

Discussiepunten. Wat is het verschil tussen “goed lesgeven” en formatief toetsen, met name als je kijkt vanuit het meer ongepland en informeel formatief toetsen in de kleine FT-cyclus waarin je vragen stelt, een eerste analyse uitvoert, klasdiscussies hebt, hierin zoekt naar zwakke punten in de klas en vervolgens nog een keer uitlegt?

Wat is de relatie tussen deze manier van formatief toetsen (de meer informele, ongeplande manier van formatief toetsen) en summatief toetsen? Concrete vragen hierbij kwamen voornamelijk vanuit het VO en waren 1) Kun je deze manier inzetten om het aantal summatieve toetsen te verminderen, en 2) hoe ga je om met vastgelegde, methodegebonden summatieve toetsen die niet passen bij formatief toetsen gericht op begrip, inzicht en student-sturing? Dit leidde tot interessante discussies over de professionaliteit van de docent om de student voortdurend te monitoren en het vertrouwen dat de docent hierin moet krijgen, mits hij/zij voldoende bekwaamheid heeft, en het belang van alignment tussen datgene wat je in formatieve toetsen nastreeft en wat in summatieve toetsen getoetst wordt.

Een aanpalend discussiepunt was: Hoe verhouden de bevindingen van deze reviewstudie, die sterk gericht zijn op formatieve toetsgedragingen in de klas, zich tot formatieve toetspraktijken op programma/curriculum niveau?

Een laatste discussiepunt was: Pleit deze overzichtsstudie nu tot professionalisering vanuit kleine experimentjes door docenten in hun eigen klas in plaats van professionalisering vanuit een school centraal geformuleerde nieuwe visie op toetsen? (discussiepunt ingebracht door beleidsadviseur)

Hoofdstuk 6: Conclusies en discussie

Dit hoofdstuk vat de belangrijkste conclusies en nieuwe inzichten van deze studie samen per onderzoeksvraag.

Onderzoeksvraag 1 beschrijft de kern van deze overzichtsstudie, is het meest uitgebreid en bevat de meeste concrete voorbeelden die inzicht geven in wat de meer en minder effectieve docent DOET in de klas wanneer hij/zij een formatieve toetspraktijk realiseert. De vijf fasen van de formatieve toetscyclus (figuur 3) vormen het kader voor het beantwoorden van deze vraag. Een eerste belangrijke conclusie is dat deze formatieve toetscyclus met de vijf fasen in geen enkele studie in zijn totaliteit is teruggekomen, terwijl verschillende studies wel het belang van de samenhang (alignment) tussen verschillende fasen aantonen, in termen van meer effectieve formatieve toetspraktijken (bijvoorbeeld, doelgericht gekozen toetsmethode (fase 1-2); feedback gekoppeld aan leerdoelen (fase 1-4), of vervolgstategie gekozen op basis van eerder geïdentificeerde zwakheden (fase 3-5)). Een formatieve toetscyclus waar wel veel studies over gevonden zijn is de “kleine formatieve toetscyclus” (2-3-5 of 2-3-2-3→5; zie figuur 4). Verschillende studies tonen aan dat docenten in deze cyclus goed en meerdere malen doorlopen meer effectieve formatieve toetspraktijken laten zien, in sommige studies ook resulterend in hogere toetsscores van studenten. Het idee om de effectiviteit van formatieve toetspraktijken van de docent te bekijken met een cyclische bril wordt hiermee in deze studie wel bevestigd.

Kwantitatieve gegevens lieten zien de meeste concrete docentacties gevonden zijn voor fase 2 (ontlokken van studentreacties). Dit is niet verrassend omdat het hier gaat om de toetsmethoden/instrumenten die docenten inzetten in het formatieve proces. Zonder het ontlokken van studentreacties kan er überhaupt nooit formatief toetsen plaatsvinden. Er is weinig gevonden over concrete docentgedragingen met betrekking tot fase 4 (communiceren met studenten over resultaten; feedback geven) en fase 1 komt ook beduidend minder aan bod. Verder zien we dat de studies laten zien dat docenten in fase 5 (het ondernemen van vervolgstappen) relatief het vaakst NIET het gedrag vertonen dat bedoeld is (“not” coderingen). De kwalitatieve analyses onderbouwen dat docenten het minste handelingsbekwaam zijn voor deze laatste fase van de FT-toetscyclus.

Wat doen docenten in fase 1? We kunnen concluderen dat zij (1) formuleren heldere en doelgerichte leerdoelen en succescriteria en gebruiken learning progressions als voorwaarde voor formatief toetsen. Dit houdt ook in dat de doelen mede bepaald worden door het doel dat de docent met formatief toetsen heeft (bv. om kennisontwikkeling bij studenten te stimuleren of om meer autonomie te creëren); (2) communiceren leerdoelen en succescriteria naar studenten, op meerdere manieren en momenten tijdens een les/lessenreeks. Vooral ook het begin en eind van iedere les is dit belangrijk; (3) betrekken studenten actief bij het gezamenlijk expliciteren van succescriteria, en (4) weten een balans te vinden tussen helder gespecificeerde doelen en te dichtgetimmerde doelen.

Met betrekking tot fase 2 kenmerken de effectievere docenten zich door: (1) ontwerpen en inzetten van doelgerichte methodieken, gekoppeld aan leerdoelen; (2) het gebruik van een breed repertoire aan methodieken, zowel formeel een informeel, in de klas; (3) veelvuldig gebruik maken van vragen stellen en klasdiscussies gericht op het uitdiepen van begrip in plaats van zoeken naar het goede antwoord, en (4) tot stand brengen van een klassenklimaat waarin de student(en) sturing geven aan de klasdiscussies en de docent flexibel kan inspringen op wat studenten inbrengen.

Fase 3: Het analyseren en interpreteren van studentreacties komt in veel studies ofwel niet expliciet terug, of komt terug als een fase waaraan docenten (te)weinig tijd besteden en zich geregeld laten sturen door oppervlakte kenmerken als vorm/presentatie in plaats van daadwerkelijke (conceptueel) begrip van individuele studenten en de gehele klas studenten. De effectievere docent besteedt hier meer tijd aan, is gericht op het boven tafel krijgen van dieper begrip van de student en zoekt ook naar misconcepties die studenten hebben. Om hier een helder beeld van te krijgen verzamelt de docent aanvullend bewijs van de student door bijvoorbeeld door te vragen, klasdiscussies te houden of studenten te vragen elkaars antwoorden te vergelijken

en uit te diepen. Deze docenten doorlopen vaker de 2-3-2-3 cyclus alvorens ze overgaan tot een vervolgstap. Effectieve docenten zijn ook in staat om technologie ter ondersteuning van deze stappen te gebruiken.

Fase 4 gaat over het communiceren met studenten over de resultaten. Hieronder valt het geven van feedback en het aanzetten van studenten tot peer feedback. Deze reviewstudie heeft echter weinig informatie opgeleverd over wat de docent DOET in de klas in deze fase 4. Gedragingen die in deze fase gevonden zijn voor de effectievere docent zijn: 1) het geven van doelgerichte en beschrijvende feedback en adresseren van misconcepties. Hierbij gaat het vooral over de alignment tussen fase 1 en 4. Docenten die hun feedback beter en expliciet koppelden aan de doelen en de learning progressions met daarin ook misconcepties, bereikten effectievere formatieve toetspraktijken; (2) het bieden van concrete suggesties en ruimte voor verbetering. Hierbij gaat het om het expliciet inbouwen van ruimte voor de student om zich te verbeteren naar aanleiding van de feedback en deze verbetering vervolgens expliciet te waarderen. Een laatste aspect is het (3) het bieden van structuur aan zelf-assessments een peer-feedback. Dus in plaats van “ga elkaar maar feedback geven”, biedt de effectievere docent concrete handvatten voor studenten om elkaar inhoudelijk goede feedback te geven.

Het ondernemen van een vervolgstap (fase 5), zowel voor vervolgonderwijs als gericht op stimuleren van bepaald leergedrag bij studenten, blijkt de meest moeilijke fase van de FT-cyclus. Tevens is dit ook de fase die het minst expliciet in veel studies aan bod komt of uitgewerkt wordt. Vaak blijven beschrijvingen bij uitspraken als “de docent gebruikt de informatie van studenten om de les aan te passen”. Slechts een klein aantal studies beschrijft concrete en goede voorbeelden van effectieve formatieve toetspraktijken waarbij docenten 1) hun vervolgstap baseren op eerdere analyse van studentreacties waaruit zij zwakheden en misconcepties hebben gedestilleerd, in de meeste gevallen op klasniveau, en 2) hun kennis van instructiestrategieën (behorende tot hun vakdidactische kennis) kunnen inzetten om een geschikte vervolgstap te kiezen. Een enkele studie laat zien dat de docent ook in staat is om een bewuste vervolgstap in te zetten voor individuele studenten en/of de sterke studenten. De meest voorkomende praktijken laten zien dat docenten vaak een vervolgstap kiezen die niet gebaseerd is op eerdere analyses van studentreacties en dat docenten in de meeste gevallen kiezen voor ofwel herhalen van de les (reteach) of het tempo aanpassen (pacing). Met betrekking tot het ondernemen van vervolgstap gericht op het stimuleren van (ander) studentgedrag is nog minder informatie gevonden. Dit is verrassend, omdat formatief toetsen zowel in theorie als in onderwijsvisies vaak gekoppeld wordt aan het stimuleren van meer zelfgestuurd, zelfregulerend leren. De beantwoording van onderzoeksvraag 2 laat zien dat docenten die met formatief toetsen zelf-regulatie of zelfsturing willen stimuleren andere gedragingen vertonen dan docenten die formatief toetsen inzetten om bijvoorbeeld kennisontwikkeling te stimuleren. Voor een docent is het dus van belang om bij het vormgeven van effectieve formatieve toetspraktijken eerst na te gaan met welk doel een formatieve toetspraktijk in de klas ingezet wordt.

Voor onderzoeksvraag 2 is gekeken naar de relatie tussen bepaald docentgedrag in de FT-cyclus en effecten op studenten. In 42 van de 106 studies wordt iets van een studentuitkomst gemeten, waarbij de meting varieert van gestandaardiseerde kennistoetsen tot ervaringen van docenten over het effect op studenten. Vijf verschillende soorten effecten worden gevonden, te weten (1) effecten op kennis(ontwikkeling) van studenten; (2) effecten op (verbeteren van) schrijven door studenten; (3) effecten op motivatie en/of attitudes van studenten; (4) effecten op leerprocessen zoals zelf-regulatie en het zoeken van advies/feedback, en (5) effecten op (algemene) studentpercepties en/of sfeer in de klas. Hoewel op basis van de ‘metingen’ geen harde uitspraken gedaan kunnen worden over daadwerkelijke (causale) relaties tussen docentgedrag en student effecten, laat de kwalitatieve analyse wel een zeer interessante trend zien, namelijk dat afhankelijk van het effect dat docenten met de formatieve toetsing willen bereiken, hun gedrag anders is. Dit verschil is vooral zichtbaar tussen docenten die formatief toetsen gebruiken voor het stimuleren van kennisontwikkeling versus het stimuleren van zelf-regulatie/zelfsturing en autonomie. In studies waarin formatief toetsen wordt ingezet om kennisontwikkeling te stimuleren, kenmerken effectieve docenten zich door het doorlopen van de kleine

formatieve toetscyclus (2-3-5 en 2-3-2-3→5), zowel via geplande en meer formele toetsmethodieken als ook door het herhaaldelijk inzetten van meer spontane klasdiscussies gericht op het uitdiepen van conceptueel begrip, en het werken met learning progressions waarin de kennisontwikkeling en bijbehorende misconcepties een grotere rol spelen. Studies die gericht zijn op formatief toetsen om zelf-regulatie, -sturing of autonomie te stimuleren, laten docenten zien die expliciete leerdoelen formuleren op participatie en sociale norm en een rijkheid aan strategieën inzetten om studenten actief te betrekken bij de verschillende fasen van de formatieve toetscyclus.

Onderzoeksvraag 3 geeft een overzicht van wat docenten moeten kennen en kunnen om de formatieve toetspraktijken uit onderzoeksvraag 1 goed te laten zien. Hierbij zijn het kunnen toepassen van vakinhoudelijke kennis en een vakdidactisch handelingsrepertoire vele malen naar boven gekomen. Vakinhoudelijk is het zicht hebben op en kunnen werken met learning progressions en/of heldere leerdoelen een belangrijke competentie omdat dit voorwaardelijk lijkt voor het vormgeven van alle fasen in de FT-cyclus. Vakdidactisch is het vooral van belang om alignment te kunnen creëren tussen de verschillende fasen van de FT-cyclus, een rijk repertoire van assessment methodieken, zowel formeel als informeel, doelgericht en flexibel te kunnen inzetten, en om handelingsbekwaam te zijn voor het bedenken en inzetten van verschillende vervolgstategieën. Andere belangrijke competenties zijn het expliciet aan de orde stellen van de student-docent relaties en activiteiten kunnen ontwikkelen om studenten actief in te zetten in de verschillende fasen van de FT-cyclus, het kunnen gebruiken van (eenvoudige) technologie om het formatieve toetsproces te ondersteunen en het hebben van een doorleefd begrip over wat formatief toetsen is.

De laatste onderzoeksvraag ging over welke aspecten van de FT-cyclus aan bod komen in professionaliseringstrajecten en in hoeverre deze al dan niet succesvol zijn in het veranderen van docentgedrag. In 43 van de 106 studies zit een professionaliseringsinterventie opgenomen (zie Appendix 3), waarvan de meeste trajecten aandacht besteden aan de kleine formatieve toetscyclus. Veel van deze studies komen tot eenzelfde conclusie, namelijk dat docenten zich het meest ontwikkelen in het goed ontlocken van studentreacties (fase 2), kleine verbeteringen laten zien in hun interpretatie van studentreacties (fase 3), maar weinig tot niet ontwikkelen in het kiezen van geschikte vervolgstappen (fase 5). Er is duidelijk extra aandacht nodig voor het ontwikkelen van vakdidactische kennis en handelingsrepertoire om met name fase 5 beter te kunnen uitvoeren. Een andere belangrijke bevinding is dat professionaliseringsinterventies die docenten ondersteunen in het doelgericht vormgeven van twee of meerdere fasen van de FT-cyclus in samenhang (bijvoorbeeld fase 1 en 2; 1 en 4; 1 en 3, of 3 en 5) leiden tot betere formatieve toetspraktijken, dan interventies die focussen op één fase (bijvoorbeeld alleen fase 2, gericht op het inzetten van een bepaalde nieuwe methode, of fase 3 gericht op het leren analyseren van studentscores op landelijke toetsen). Een belangrijk punt is ook nog dat verschillende trajecten wel resulteerden in *meer* implementatie (bijvoorbeeld meer communiceren van doelen naar studenten), maar dat bij observatie bleek dat deze implementatie vaak van magere kwaliteit was. Twee laatste relevante conclusies uit de professionaliseringsinterventies zijn (1) dat veel interventies óók aandacht besteden aan het creëren van een doorleefd begrip van formatief toetsen, dat hier ook vaak effecten op werden gevonden, maar dat deze niet altijd direct resulteerden in zichtbare veranderingen in praktijken in de klas, en (2) dat professionalisering waarbij aandacht is voor zowel vakinhoudelijke ontwikkeling als vakdidactische ontwikkeling in de context van formatief toetsen effectief lijken te zijn, vooral voor fase 2 en 3.

Hoofdstuk 7: Limitaties en implicaties

Dit laatste hoofdstuk blikt terug en vooruit vanuit deze studie en haar resultaten. Een aantal beperkingen van de studie wordt eerst besproken, omdat deze de conclusies en implicaties in een bepaald licht zetten. Daarna beschrijven we de theoretische bijdrage met betrekking tot formatief toetsen en benoemen we praktische implicaties voor verdere professionalisering.

Limitaties

Deze studie onderscheidt zich van eerdere (review)studies naar formatief toetsen door de sterke focus op concrete gedragingen van de docent in de klas. De zoekstrategie en inclusiecriteria waren hier sterk op gericht. Hierdoor zijn de artikelen geselecteerd voor deze studie voor een groot gedeelte anders dan de beide voorgaande PPO-NRO reviews naar formatief toetsen (Schildkamp et al., 2014; Sluijsmans et al., 2013). Deze keuze heeft er ook toe geleid dat veel gevonden studies gaan over on-the-fly, spontane en kort cyclische formatieve toetspraktijken die zichtbaar zijn in de context van een les of korte lessenreeks. Andere vormen van formatief toetsen, zoals formatief toetsen op programmaniveau, komen hierdoor niet of slechts mondjesmaat aan de orde in de geselecteerde studies. Hierdoor dekt deze studie dus zeker niet alle vormen en varianten van formatief toetsen. Door deze keuze schijnt deze studie echter wel een nieuw licht op formatief toetsen, namelijk die van wat een docent concreet doet en in zijn/haar eigen lessen met zijn/haar eigen studenten.

Een aanvullende consequentie van bovenstaande focus is dat deze studie niet gaat over onderwijsvisies en opvattingen over wat goed formatief toetsen zou moeten zijn, terwijl juist veel (review)studies naar formatief toetsen laten zien dat deze aspecten voorwaardelijk zijn voor het kunnen bereiken van effectieve formatieve toetspraktijken (Sluijsmans, Joosten-Ten Brinke, & Schilt-Mol, 2014; Gulikers, Biemans, Wesselink, & van der Wel, 2013). Wij willen met deze studie zeker niet zeggen deze aspecten niet belangrijk zijn voor effectieve formatieve toetspraktijken, maar uit eerdere reviewstudies blijkt juist ook het belang van de concrete implementatie in de klas door docenten. In de professionaliseringstrajecten beschreven bij onderzoeksvraag 4 blijkt dat het werken aan het ontwikkelen van een doorleefd begrip van formatief toetsen ook vaak geïntegreerd gebeurt met het werken aan concrete formatieve toetsgedragingen. We hopen met deze studie iedere individuele docent te inspireren tot het experimenteren met formatieve toetsgedragingen in de context van de eigen les, klas en met eigen studenten

Opvallend in deze studie is dat er maar weinig studies gevonden zijn die concrete informatie gaven over wat de docent doet in fase 4 (communiceren over resultaten met studenten), terwijl deze fase over feedback gaat. Feedback wordt in vele andere studies juist vaak geduidt als de kern van formatief toetsen (zie ook de eerdere reviews van Schildkamp et al., 2014; Sluijsmans et al., 2013). Een eerste exploratieve zoektocht in databases resulteerde in heel veel studies over feedback, echter een groot deel hiervan ging over theoretische beschrijvingen van goede feedback, student percepties van feedback of experimentele studies waarin studenten verschillende varianten van vaak geschreven feedback kregen, in tegenstelling tot informatie over concrete docentfeedback praktijken in de klas. De combinatie tussen de centrale zoekterm “assessment practices” en “feedback” bleek veel minder voor te komen, omdat veel studies over feedback sec over feedback gaan en niet over het formatieve proces van feed-up-feedback-feedforward. Daarom hebben wij uiteindelijk, in navolging van de review van Kingston & Nash (2011) ervoor gekozen “feedback” niet te gebruiken als apart zoekcriterium. Dit kan een verklaring zijn voor het vinden van weinig resultaten met betrekking tot fase 4.

De geselecteerde studies in deze reviewstudie waren voor een groot deel uitgevoerd in de Verenigde Staten, in de context van primair of voortgezet onderwijs en in de context van een duidelijk schoolvak (science, math, reading/writing). Dit kan de generaliseerbaarheid van de bevindingen beïnvloeden. Zo was er in deze studies vaak sprake van een nationaal (of state-wide) curriculum en kan het wellicht zo zijn dat in deze duidelijk vakgerichte studies op primair/voortgezet onderwijs het maken van een learning progression makkelijker

mogelijk is, dan bijvoorbeeld voor een beroepsopleiding. Desalniettemin laten de reacties van een breed scala aan docenten tijdens de kennisdeeldag zien dat de resultaten herkend worden en men juist geïnspireerd raakte door ideeën over bijvoorbeeld learning progressions en misconcepties. Wel moeten we kijken naar of en hoe de resultaten in de Nederlandse context en op verschillende lagen van het onderwijs concrete invulling kunnen krijgen.

Voor onderzoeksvraag 2 is gekeken naar de effecten van concrete formatieve toetspraktijken van docenten op studenten. Hierbij zijn de effectmetingen veelal gebaseerd op zwakke onderzoeksdesigns en docentervaringen. De gevonden relaties mogen dan ook zeker niet als causaal geïnterpreteerd worden. Echter, de gevonden trend dat effectief formatief toetsgedrag van docenten anders is wanneer formatief toetsen wordt ingezet voor kennisontwikkeling dan wel voor het stimuleren van zelfregulerend leren is een interessante bevinding die verder onderzoek mogelijk maakt en tevens benadrukt hoe belangrijk het is om eerst te bedenken voor welk doel je formatief gaat toetsen alvorens te bepalen welk soort van gedragingen en praktijken hiervoor noodzakelijk zijn.

Voor het destilleren van noodzakelijke formatieve toetscompetenties (onderzoeksvraag 3) kijkt deze studie puur vanuit de effectieve docentgedragingen (onderzoeksvraag 1). Er is geen aparte zoekactie uitgevoerd naar formatieve toetscompetenties van docenten. Wij willen dan ook niet pretenderen dat de set van geïdentificeerde toetscompetenties volledig is. Echter, competentiebeschrijvingen hebben de neiging om vaag en algemeen te zijn en krijgen pas echt betekening in een bepaalde context en taak (Mulder, 2014). Door vanuit concreet docentgedrag te kijken naar wat docenten moeten kennen en kunnen om dit gedrag te laten zien, wordt de relatie tussen de benodigde formatieve toetscompetenties en daadwerkelijk gedrag concreter, wat ook kan leiden tot concretere ideeën voor verdere, doelgerichte professionalisering.

Theoretische implicaties

De cyclische benadering van formatief toetsen via de vijf fasen, de focus op concrete docentpraktijken in de klas en de koppeling hiervan aan effecten op studenten door te kijken naar studenteffecten op allerlei manieren ‘gemeten’ leveren nieuwe theoretische inzichten op voor verdere theorie over en onderzoek naar formatief toetsen.

Het doel van deze reviewstudie was om de theorie en praktijk van formatief toetsen een stap verder te helpen door gedetailleerd en concreet te kijken naar wat docenten doen in hun klas als ze een formatieve toetspraktijk realiseren. Dit sluit aan bij de roep van recente studies (Antoniou & James, 2014) om formatief toetsonderzoek meer te verplaatsen naar het in kaart brengen van daadwerkelijke docentpraktijken in de klas. De formatieve toetscyclus, en haar vijf fasen, gebruikt in deze studie heeft uitgewezen handzaam te zijn om deze concrete docentpraktijken te duiden en te differentiëren. In deze FT-cyclus zijn van meet af aan de “assessment for learning stroming” en de “data-based decision making” stroming met elkaar verbonden (zie hoofdstuk 2). Gedurende deze reviewstudie, kwamen de vijf fasen in verschillende samenstellingen in verschillende artikelen terug. Een interessante bevinding hierbij was dat deze bijdragen niet altijd hun origine hadden in formatief toetsonderzoek of praktijk, maar juist kwamen uit meer instructie-georiënteerde stromingen, zoals scaffolding (van der Pol et al., 2013; 2014; Pat-el et al., 2012), inquiry teaching (Rice & Smith, 20110); responsive instruction (Sabel et al., 2015), of noticing (Barnhart & Van Es, 2015). Door de centrale focus van deze review op docentpraktijken in de klas, en de nauwe verbondenheid tussen formatief toetsen en goede instructie, kwamen in de geselecteerde artikelen ook deze meer instructie-georiënteerde artikelen terug.

Vervolgonderzoek naar formatief toetsen kan verrijkt worden door expliciet aanknopingspunten met dit soort aanpalende onderzoeksdomeneinen aan te gaan. Zo is bij formatief toetsen de doelgerichtheid een belangrijk en verder uitgewerkt aspect, terwijl de instructie-georiënteerde modellen vaak gedetailleerder ingaan op het concrete docent-student interactieproces. De kracht van dergelijke combinaties werd recentelijk ook beargumenteerd op de AERA waar Furtak, Thompson en van Es (2016) twee van deze stromingen integreerden, namelijk formatief toetsen en noticing (i.e., the process by which teachers see and make sense of particular

events during classroom instruction, Furtak et al., 2016, p.1). Deze integratie leidde tot een rijker analysekader om naar docentpraktijken in de klas te kijken en daarmee beter grip te krijgen op het complexe proces van formatief toetsen en hiermee ook het onderscheid met “goede instructie” te verhelderen.

Wat betreft het effect op studenten (onderzoeksvraag 2) laat deze studie zien dat het verbreden van de blik op studenteffecten, verder dan experimentele pre-post test studies waardevolle informatie oplevert. Er is nu eenmaal weinig experimenteel onderzoek naar de effectiviteit van formatief toetsen en de studies die er zijn, zijn veelal beperkt tot het meten van makkelijk meetbare effecten via bijvoorbeeld gestandaardiseerde kennistoetsen, terwijl we met formatief toetsen juist vaak andere doelen dan kennisontwikkeling nastreven. In deze studie hebben we in principe geen beperkingen opgelegd aan hoe studenteffecten gemeten zijn. Natuurlijk onderkennen we dat door deze keuze meerdere studies in deze review methodologisch zwakkere onderzoeksdesigns hebben, maar juist door al deze verschillende studies mee te nemen, die ook verschillende leereffecten of leerprocessen nastreven, laat deze studie zien dat docentgedragingen inderdaad verschillend lijken te zijn afhankelijk van het doel (leeruitkomst/leerproces) dat wordt beoogd met formatief toetsen. De resultaten bij onderzoeksvraag 2 lijken duidelijke verschillen te laten zien tussen docentgedragingen die kennisontwikkeling respectievelijk zelf-regulatie en autonomie stimuleren. De FT-cyclus ziet er dus wellicht anders uit, in ieder geval wat betreft docentgedragingen en student-docent interactie, afhankelijk van het doel dat de docent met formatief toetsen nastreeft. Daarmee bevestigt deze studie de cruciale rol van de student-docent interactie voor formatief toetsen (zie ook bv. Buck et al., 2010). Vervolgonderzoek kan de gevonden trends tussen docentgedrag en bepaalde studenteffecten verder en specifiekere onderzoeken, met hierbij ook aandacht voor de specifieke student-docent interacties, en daarmee meer empirisch onderbouwd inzicht bieden in welke formatieve gedragingen meer of minder effectief zijn voor het stimuleren van verschillende effecten op studenten.

Praktische implicaties: Suggesties voor professionalisering

Uit deze reviewstudie komen een aantal nieuwe inzichten naar voren voor verdere professionalisering van docenten in formatief toetsen.

- Aandacht voor de gehele FT-cyclus en het creëren van doelbewuste samenhang (alignment) tussen de fasen
- Aandacht voor de doelen/ beoogde effecten op studenten waarvoor formatief toetsen ingezet kan worden en dat afhankelijk van de effecten die docenten willen bereiken andere FT-gedragingen van docenten nodig zijn in de FT-cyclus.
- Werken met learning progressions en misconcepties en hier gebruik van maken in de verschillende FT-fasen
- Ideeën opdoen voor en experimenteren met hoe fase 3 (het analyseren en interpreteren van studentreacties) vorm kan krijgen, zonder dat dit te veel tijd kost, een actief onderdeel van de les kan zijn, en goed en eenvoudig inzicht geeft in zwakheden, misconcepties en sterke punten van individuele leerlingen, groepen en hele klassen. Bijvoorbeeld, een effectieve strategie en direct zinvolle activiteit in de klas is ‘clusteren en categoriseren’ samen met studenten, terwijl deze zelfs door de meest effectieve docenten maar mondjesmaat wordt toegepast (Ruiz-Primo & Furtak, 2007).
- Oefenen met allerlei mogelijkheden om de student-docent relatie expliciet aan de orde te stellen en studenten op verschillende manieren actief te maken in het expliciteren en uitdiepen van hun eigen begrip en leerproces in de verschillende fasen van de FT-cyclus.
- Inzetten op het ontwikkelen van een breder handelingsrepertoire en handelingsbekwaamheid voor het vormgeven van vervolgstategieën. Deze aanbeveling sluit nauw aan bij de sterke focus die er in Nederland de laatste jaren meer en meer gekomen is, mede doordat de onderwijsinspectie hier sterk op gericht is, op het differentiëren in de klas. De kennisdeelnijvenkomst bevestigde dat docenten handelingsverlegen zijn als het gaat om het kiezen van verschillende vervolgstappen en bij differentiëren het gevoel hebben iedere individuele student aandacht te moeten geven, wat zij vervolgens als onmogelijk ervaren gezien de grote

groepen en beperkte tijd. Het denken in ‘massamaatwerk” waarbij op basis van analyses van studentreacties groepen gevormd worden die de docent vervolgens als groepen gedifferentieerd bediend, verdient hierbij veel aandacht.

Referenties

Referenties meegenomen in de review (N= 106)

- Abrams, L. M., & McMillan, J. H. (2013). The instructional influence of interim assessments: Voices from the field. R. W. Lissitz (Ed). *Informing the practice of teaching using formative and interim assessment: A systems approach (Chapter 5)*. Charlotte, NC: Information Age Publishing.
- Antoniou, P., & James, M. (2014). Exploring formative assessment in primary school classrooms: Developing a framework of actions and strategies. *Educational Assessment, Evaluation and Accountability*, 26(2), 153-176.
- Aschbacher, P., & Alonzo, A. (2006). Examining the utility of elementary science notebooks for formative assessment purposes. *Educational Assessment*, 11(3-4), 179-203.
- Ateh, C. M. (2015). Science teachers' elicitation practices: Insights for formative assessment. *Educational Assessment*, 20(2), 112-131
- Ayala, C. C., Shavelson, R. J., Araceli Ruiz-Primo, M., Brandon, P. R., Yin, Y., Furtak, E. M., et al. (2008). From formal embedded assessments to reflective lessons: The development of formative assessment studies. *Applied Measurement in Education*, 21(4), 315-334.
- Bailey, A. L., & Heritage, M. (2014). The role of language learning progressions in improved instruction and assessment of English language learners. *TESOL Quarterly*, 48(3), 480-506
- Barnhart, T., & van Es, E. (2015). Studying teacher noticing: Examining the relationship among pre-service science teachers' ability to attend, analyze and respond to student thinking. *Teaching and Teacher Education*, 45 Jan, 83-93.
- Beesley, A. (2009). Measuring classroom assessment with a work sample. *Online Submission*
- Bloxham, S., & Campbell, L. (2010). Generating dialogue in assessment feedback: Exploring the use of interactive cover sheets. *Assessment & Evaluation in Higher Education*, 35(3), 291-300.
- Brookhart, S. M., Moss, C. M., & Long, B. A. (2010). Teacher inquiry into formative assessment practices in remedial reading classrooms. *Assessment in Education: Principles, Policy & Practice*, 17(1), 41-58.
- Buck, G. A., TrauthNare, A., & Kaftan, J. (Apr 2010). Making formative assessment discernable to pre-service teachers of science. *Journal of Research in Science Teaching*, 47(4), 402-421
- Chen, Q., May, L., Klenowski, V., & Kettle, M. (Jul 2014). The enactment of formative assessment in English language classrooms in two Chinese universities: Teacher and student responses. *Assessment in Education: Principles, Policy & Practice*, 21(3), 271-285
- Chin, C., & Teou, L. (2010). Formative assessment: Using concept cartoon, pupils' drawings, and groupdiscussions to tackle children's ideas about biological inheritance. *Biological Education*, 44(3), 108-115.
- Clark, I. (Jun 2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, 24(2), 205-249
- Conway, R. (2011). Owning their learning: Using 'assessment for learning' to help students assume responsibility for planning, (some) teaching and evaluation. *Teaching History*, (144), 51.
- Davis, D. S., & Neitzel, C. (2011). A self-regulated learning perspective on middle grades classroom assessment. *The Journal of Educational Research*, 104(3), 202-215.
- De Lisle, J. (2015). The promise and reality of formative assessment practice in a continuous assessment scheme: The case of Trinidad and Tobago. *Assessment in Education: Principles, Policy & Practice*, 22(1), 79-103
- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence on the impact of formative assessment in education. *Practical Assessment, Research & Evaluation*, 14, 1-11.
- Falk, B., Ort, S. W., & Moirs, K. (2007). Keeping the focus on the child: Supporting and reporting on teaching and learning with a classroom-based performance assessment system. *Educational Assessment*, 12(1), 47-75
- Falk, A. (Mar 2012). Teachers learning from professional development in elementary science: Reciprocal relations between formative assessment and pedagogical content knowledge. *Science Education*, 96(2), 265-290.
- Feldman, A., & Capobianco, B. M. (2008). Teacher learning of technology enhanced formative assessment. *Journal of Science Education and Technology*, 17(1), 82-99.

- Fluckiger, J., Vigil, Y. T. y., Pasco, R., & Danielson, K. (2010). Formative feedback: Involving students as partners in assessment to enhance learning. *College Teaching*, 58(4), 136-140.
- Forbes, C. T., Sabel, J. L., & Biggers, M. (2015). Elementary teachers' use of formative assessment to support students' learning about interactions between the hydrosphere and geosphere. *Journal of Geoscience Education*, 63(3), 210-221.
- Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., & Katzaroff, M. (1999). Mathematics performance assessment in the classroom: Effects on teacher planning and student problem solving. *American Educational Research Journal*, 36(3), 609-646.
- Furtak, E. M. (2012). Linking a learning progression for natural selection to teachers' enactment of formative assessment. *Journal of Research in Science Teaching*, 49(9), 1181-1210.
- Furtak, E. M., & Heredia, S. C. (Oct 2014). Exploring the influence of learning progressions in two teacher communities. *Journal of Research in Science Teaching*, 51(8), 982-1020.
- Furtak, E. M., & Ruiz-Primo, M. A. (2008). Making students' thinking explicit in writing and discussion: An analysis of formative assessment prompts. *Science Education*, 92(5), 799-824.
- Furtak, E. M., Ruiz-Primo, M. A., Shemwell, J. T., Ayala, C. C., Brandon, P. R., Shavelson, R. J., et al. (2008). On the fidelity of implementing embedded formative assessments and its relation to student learning. *Applied Measurement in Education*, 21(4), 360-389.
- Gearhart, M., Nagashima, S., Pfothner, J., Clark, S., Schwab, C., Vendlinski, T., et al. (2006). Developing expertise with classroom assessment in K-12 science: Learning to interpret student work. interim findings from a 2-year study. *Educational Assessment*, 11(3-4), 237-263.
- Gearhart, M., & Osmundson, E. (2009). Assessment portfolios as opportunities for teacher learning. *Educational Assessment*, 14(1), 1-24.
- Gerard, L. F., Spitulnik, M., & Linn, M. C. (Nov 2010). Teacher use of evidence to customize inquiry science instruction. *Journal of Research in Science Teaching*, 47(9), 1037-1063.
- Goertz, M. E., Olah, L. N., & Riggan, M. (2009). Can interim assessments be used for instructional change? Policy brief. RB-51. *Consortium for Policy Research in Education*.
- Gómez, M. C., & Jakobsson, A. (2014). Everyday classroom assessment practices in science classrooms in Sweden. *Cultural Studies of Science Education*, 9(4), 825-853.
- Gonzales, Richard D. L. C., & Fuggan, C. G. (Jan 2012). Exploring the conceptual and psychometric properties of classroom assessment. *The International Journal of Educational and Psychological Assessment*, 9(2), 45-60.
- Gottheiner, D. M., & Siegel, M. A. (2012). Experienced middle school science teachers' assessment literacy: Investigating knowledge of students' conceptions in genetics and ways to shape instruction. *Journal of Science Teacher Education*, 23(5), 531-557.
- Graney, S. B., & Shinn, M. R. (2005). Effects of reading curriculum-based measurement (R-CBM) teacher feedback in general education classrooms. *School Psychology Review*, 34(2), 184.
- Hawe, E. M., & Dixon, H. R. (2014). Building students' evaluative and productive expertise in the writing classroom. *Assessing Writing*, 19, 66-79.
- Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (Mar 2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. *Journal of Research in Science Teaching*, 49(3), 333-362.
- Herman, J., Osmundson, E., Dai, Y., Ringstaff, C., & Timms, M. (2015). Investigating the dynamics of formative assessment: Relationships between teacher knowledge, assessment practice and learning. *Assessment in Education: Principles, Policy and Practice*, 22(3), 344-367.
- Herpich, S., Wittwer, J., Nuckles, M., & Renkl, A. (Nov 2014). Addressing knowledge deficits in tutoring and the role of teaching experience: Benefits for learning and summative assessment. *Journal of Educational Psychology*, 106(4), 934-945.
- Hickey, D. T., & Anderson, K. T. (2007). Situative Approaches to Student Assessment: Contextualizing Evidence to Transform Practice. *Yearbook of the National Society for the Study of Education* 106(1), 264-287.
- Hogan, D. A. V. I. D., Towndrow, P., & Koh, K. (2009). Instructional and assessment practices in Singapore. Grigorenko, E. L. (Ed.). *Multicultural psychoeducational assessment* (Chapter 9). Springer Publishing Company.
- Jones, A., & Moreland, J. (2005). The importance of pedagogical content knowledge in assessment for learning practices: A case-study of a whole-school approach. *Curriculum Journal*, 16(2), 193-206.
- Kang, H., Thompson, J., & Windschitl, M. (Jul 2014). Creating opportunities for students to show what they know: The role of scaffolding in assessment tasks. *Science Education*, 98(4), 674-704.

- Kay, R., & Knaack, L. (2009). Exploring the use of audience response systems in secondary school science classrooms. *Journal of Science Education and Technology*, 18(5), 382-392.
- Kearney, S. (Nov 2013). Improving engagement: The use of 'authentic self-and peer-assessment for learning' to enhance the student learning experience. *Assessment & Evaluation in Higher Education*, 38(7), 875-891.
- Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education*, 112(4), 496-520.
- Kingston, N., & Nash, B. (Win 2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28-37.
- Lee, I. (2011). Formative assessment in EFL writing: An exploratory case study. *Changing English: Studies in Culture and Education*, 18(1), 99-111.
- Lorente, E., & Kirk, D. (Jan 2013). Alternative democratic assessment in PETE: An action-research study exploring risks, challenges and solutions. *Sport, Education and Society*, 18(1), 77-96.
- Lyon, E. G. (2011). Beliefs, practices, and reflection: Exploring a science teacher's classroom assessment through the assessment triangle model. *Journal of Science Teacher Education*, 22(5), 417-435.
- Lyon, E. G. (2013). Learning to assess science in linguistically diverse classrooms: Tracking growth in secondary science preservice teachers' assessment expertise. *Science Education*, 97(3), 442-467.
- Lysaght, Z., & O'Leary, M. (2013). An instrument to audit teachers' use of assessment for learning. *Irish Educational Studies*, 32(2), 217-232.
- Matuk, C. F., Linn, M. C., & Eylon, B. (2015). Technology to support teachers using evidence from student work to customize technology-enhanced inquiry units. *Instructional Science*, 43(2), 229-257.
- McDowell, L., Wakelin, D., Montgomery, C., & King, S. (2011). Does assessment for learning make a difference? the development of a questionnaire to explore the student response. *Assessment & Evaluation in Higher Education*, 36(7), 749-765.
- McNaughton, S., Lai, M. K., & Hsiao, S. (2012). Testing the effectiveness of an intervention model based on data use: A replication series across clusters of schools. *School Effectiveness and School Improvement*, 23(2), 203-228.
- McNeill, M., Gosper, M., & Xu, J. (2012). Assessment choices to target higher order learning outcomes: The power of academic empowerment. *Research in Learning Technology*, 20(3), 283-296.
- MittonKukner, J., & Orr, A. M. (Sum 2014). Making the invisible of learning visible: Pre-service teachers identify connections between the use of literacy strategies and their content area assessment practices. *Alberta Journal of Educational Research*, 60(2), 403-419.
- Moss, C. M., Brookhart, S. M., & Long, B. A. (Jul 2013). Administrators' roles in helping teachers use formative assessment information. *Applied Measurement in Education*, 26(3), 205-218.
- Mui So, W. W., & Hoi Lee, T. T. (2011). Influence of teachers' perceptions of teaching and learning on the implementation of assessment for learning in inquiry study. *Assessment in Education: Principles, Policy & Practice*, 18(4), 417-432.
- Mulder, M. (2014). Conceptions of Professional Competence. S. Billett, C. Harteis, & H. Gruber (Eds). *International Handbook of Research in Professional and Practice-based Learning* (p. 107-137). Dordrecht: Springer.
- Newby, L., & Winterbottom, M. (2011). Can research homework provide a vehicle for assessment for learning in science lessons? *Educational Review*, 63(3), 275-290.
- Chróinín, D. N., & Cosgrave, C. (2013). Implementing formative assessment in primary physical education: Teacher perspectives and experiences. *Physical Education and Sport Pedagogy*, 18(2), 219-233.
- Offerdahl, E., & Tomanek, D. (2011). Changes in instructors' assessment thinking related to experimentation with new strategies. *Assessment & Evaluation in Higher Education*, 36(7), 781-795.
- O'Loughlin, J., Chróinín, D. N., & O'Grady, D. (2013). Digital video: The impact on children's learning experiences in primary physical education. *European Physical Education Review*, 19(2), 165-182.
- Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, 22(6), 806-813.
- Panizzon, D., & Pegg, J. (2008). Assessment practices: Empowering mathematics and science teachers in rural secondary schools to enhance student learning. *International Journal of Science and Mathematics Education*, 6(2), 417-436. Retrieved from SCOPUS database.
- Parr, J. M., & Limbrick, L. (Apr 2010). Contextualising practice: Hallmarks of effective teachers of writing. *Teaching and Teacher Education*, 26(3), 583-590.
- Pat-El, R., Tillema, H., & van Koppen, S. W. (2012). Effects of formative feedback on intrinsic motivation: Examining ethnic differences. *Learning and Individual Differences*, 22(4), 449-454.

- Penuel, W. R., Boscardin, C. K., Masyn, K., & Crawford, V. M. (2007). Teaching with student response systems in elementary and secondary education settings: A survey study. *Educational Technology Research and Development*, 55(4), 315-346.
- Pham, T. H. T., & Renshaw, P. (Jan 2015). Formative assessment in confucian heritage culture classrooms: Activity theory analysis of tensions, contradictions and hybrid practices. *Assessment & Evaluation in Higher Education*, 40(1), 45-59.
- Phelan, J. C., Choi, K., Niemi, D., Vendlinski, T. P., Baker, E. L., & Herman, J. (2012). The effects of POWERSOURCE© assessments on middle-school students' math performance. *Assessment in Education: Principles, Policy & Practice*, 19(2), 211-230.
- Poskitt, J. (2014). Transforming professional learning and practice in assessment for learning. *Curriculum Journal*, 25(4), 542-566.
- Restrepo, H. N. A. (2013). Role of systematic formative assessment on students' views of their learning. *PROFILE Issues in Teachers' Professional Development*, 15(2), 165-183.
- Rice, K., & Smith, R. (2011). Innovative ways to use data to inform planning of teaching to enhance student learning. *Inted2011: 5th International Technology, Education and Development Conference*, , 4955-4964.
- Riggan, M., & Olah, L. N. (Jan 2011). Locating interim assessments within teachers' assessment practice. *Educational Assessment*, 16(1), 1-14.
- Roberts, T. S. (Ed.). (2006). *Self, peer and group assessment in e-learning*. IGI Global.
- Robinson, J., Myran, S., Strauss, R., & Reed, W. (2014). The impact of an alternative professional development model on teacher practices in formative assessment and student learning. *Teacher Development*, 18(2), 141-162.
- Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation*, 37(1), 15-24.
- Ruiz-Primo, M. A., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment*, 11(3-4), 205-235.
- Ruiz-Primo, M. A., & Furtak, E. M. (Jan 2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44(1), 57-84.
- Sabel, J. L., Forbes, C. T., & Zangori, L. (2015). Promoting prospective elementary teachers' learning to use formative assessment for life science instruction. *Journal of Science Teacher Education*, 26(4), 419-445.
- Sato, M., Wei, R. C., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of national board certification. *American Educational Research Journal*, 45(3), 669-700.
- Schneider, M. C., & Andrade, H. (Jul 2013). Teachers' and administrators' use of evidence of student learning to take action: Conclusions drawn from a special issue on formative assessment. *Applied Measurement in Education*, 26(3), 159-162.
- Shapiro, E. S. (2011). *Academic skills problems: Direct assessment and intervention*. New York, NY, US: Guilford Press
- Shepard, L., Hammerness, K., DarlingHammond, L., Rust, F., BaratzSnowden, J., Gordon, E., et al. (2005). Chapter: Assessment., 275-326.
- Shirley, M. L., & Irving, K. E. (Feb 2015). Connected classroom technology facilitates multiple components of formative assessment practice. *Journal of Science Education and Technology*, 24(1), 56-68.
- Simpson-Beck, V. (2011). Assessing classroom assessment techniques. *Active Learning in Higher Education*, 12(2), 125-132.
- Strowbridge, J. (2009). Middle school teachers' use of a formative feedback guide in mathematics problem solving instruction. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 69(7-A), 2646.
- Suurtamm, C., & Koch, M. J. (2014). Navigating dilemmas in transforming assessment practices: Experiences of mathematics teachers in Ontario, Canada. *Educational Assessment, Evaluation and Accountability*, 26(3), 263-287.
- Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy & Practice*, 17(4), 399-417.
- Talanquer, V., Bolger, M., & Tomanek, D. (May 2015). Exploring prospective teachers' assessment practices: Noticing and interpreting student understanding in the assessment of written work. *Journal of Research in Science Teaching*, 52(5), 585-609.
- Talanquer, V., Tomanek, D., & Novodvorsky, I. (Feb 2013). Assessing students' understanding of inquiry: What do prospective science teachers notice? *Journal of Research in Science Teaching*, 50(2), 189-208.

- Tan, K. H. K., & Prosser, M. (2004). Qualitatively different ways of differentiating student achievement: A phenomenographic study of academics' conceptions of grade descriptors. *Assessment & Evaluation in Higher Education*, 29(3), 267-282.
- Tang, S. Y. F., Leung, P. P. W., Chow, A. W. K., & Wong, P. M. (2010). A case study of teacher learning in an assessment for learning project in Hong Kong. *Professional Development in Education*, 36(4), 621-636.
- Torrance, H. (2007). Assessment as learning? how the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. 1. *Assessment in Education*, 14(3), 281-294.
- Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal*, 27(5), 615-631.
- TrauthNare, A., & Buck, G. (Sep 2011). Using reflective practice to incorporate formative assessment in a middle school science classroom: A participatory action research study. *Educational Action Research*, 19(3), 379-398.
- van de Pol, J., Volman, M., Oort, F., & Beishuizen, J. (2014). Teacher scaffolding in small-group work: An intervention study. *Journal of the Learning Sciences*, 23(4), 600-650.
- van de Pol, J., Volman, M., & Beishuizen, J. (Feb 2012). Promoting teacher scaffolding in small-group work: A contingency perspective. *Teaching and Teacher Education*, 28(2), 193-205.
- Wall, P. T. C. (2012). Using common formative assessments to promote student achievement: A case study of practice, leadership, and culture. *ProQuest LLC.Ed.D.Dissertation, University of Missouri - Columbia*.
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice*, 11(1), 49-65.
- Willis, J. (2011). Affiliation, autonomy and assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 18(4), 399-415.
- Wingate, U. (2010). The impact of formative feedback on the development of academic writing. *Assessment & Evaluation in Higher Education*, 35(5), 519-533.
- Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: Issues of breadth and quality. *Assessment in Education: Principles, Policy & Practice*, 22(1), 140-160.
- Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M., Brandon, P. R., Furtak, E. M., et al. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education*, 21(4), 335-359.
- Yin, Y., Tomita, M. K., & Shavelson, R. J. (2014). Using formal embedded formative assessments aligned with a short-term learning progression to promote conceptual change and achievement in science. *International Journal of Science Education*, 36(4), 531-552.

Overige referenties

- Baartman, L., & Gulikers, J. (2014). Beoordeling als fundament van goed beroepsonderwijs: een analyse van 11 toetsprogramma's in het mbo en hbo. *Pedagogische Studiën*, 91, 54-68.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25.
- Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, 24(2), 205-249.
- Coffey, J. E., Hammer, D., Levin, D. M., & Grant, T. (2011). The missing disciplinary substance of formative assessment. *Journal of Research in Science Teaching*, 48(10), 1109-1136. doi: 10.1002/tea.20440
- Furtak, E. M., Thompson, J., & van Es, E. (2016, April) Formative Assessment and Noticing: Toward a Synthesized Framework for Attending and Responding During Instruction. Paper Presented at the Annual Meeting of the American Educational Research Association, Washington, D.C., USA.
- Gulikers, J. T. M., Biemans, H. J. A., Wesselink, R., & van der Wel, M. (2013). Aligning formative and summative assessments: A collaborative action research challenging teacher conceptions. *Studies in Educational Evaluation*, 39(2), 116-124. doi: <http://dx.doi.org/10.1016/j.stueduc.2013.03.001>
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81-112. doi: 10.3102/003465430298487
- Inspectie van het Onderwijs (2011). Opbrengsten maak er meer van! Retrieved from: www.onderwijsinspectie.nl
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational measurement: Issues and practice*, 30(4), 28-37.

- Kleijn, R. A. M. de, Prins F. J., Lutz, Ch., van Look, K. C. J., van Tartwijk, J. (2015). *Impact of assessment on motivation. Effects in different learning contexts*. NRO review 411-12-225. NRO: Den Haag.
- Lai, M. K. & Schildkamp, K. (2014). Data-based decision making: an overview. In K. Schildkamp, M. K. Lai, & Earl, L. (Eds). *Data-based Decision Making in Education*. Dordrecht: Springer
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Oxford, UK: Blackwell.
- Schildkamp, K., Lai, M. K., & Earl, L. (Eds.) (2014), *Data-based Decision Making in Education* (Vol. 17). Dordrecht: Springer Netherlands.
- Schildkamp, K., Heitink, M., van der Kleij, F., Hoogland, I., Dijkstra, A., Kippers, W. & Veldkamp, B. (2014). *Voorwaarden voor effectieve formatieve toetsing. Een praktische review*. PPO-NRO 405-14-534. Den Haag: NRO
- Sluijsmans, D., Joosten-Ten Brinke, D. & Schilt-Mol, T. (2015). *De kwaliteit van toetsing onder de loep. handvatten om de kwaliteit van toetsing in het hoger onderwijs te analyseren, verbeteren en borgen*. Apeldoorn: Garant Uitgevers
- Sluijsmans, D. M. A., Joosten-ten Brinke, D., & Van der Vleuten, C. P. M. (2013). *Toetsen met leerwaarde. Een reviewstudie naar effectieve kenmerken van formatief toetsen*. NWO-PROO 411-11-697. Den Haag: NWO.
- Van der Kleij, F. M., Vermeulen, J. A., Schildkamp, K., & Eggen, T. J. H. M. (2015). Integrating data-based decision making, Assessment for Learning and diagnostic testing in formative assessment. *Assessment in Education: Principles, Policy & Practice*, 22(3), 324-343. doi: 10.1080/0969594x.2014.999024
- Vermaas, J. (2011). *Opbrengst gericht werken en het vakmanschap van de leraar*. Den Haag: School aan zet.
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14.

Appendices

Appendix 1. Codeerschema voor codering in Atlas-Ti

Overall categories (for the whole article)

educational level:

- primary
- secondary
- higher education
- teacher education

Educational field

- Science
- language
- Math
- other field

Country

- USA
- UK
- European country, not UK
- Singapore/Asia
- Other country

Professionalisation as intervention (if yes, than code)

Learning outcome measured? (if yes, then code)

Research methodology

- quantitative
- qualitative
- Mixed methodology
-

FA phases explicitly addressed

- Phase 1
- Phase 2
- Phase 3
- Phase 4
- Phase 5

Coding withing the paper:

The following codes should just be put on the part of the paper describing this topic

- Definition of FA
- Research question(s)
- Measured learning outcome(s)
- Description of Professionalization approach
- Description of methodology
- Data analysis
- Main conclusions
- Effect of professional development intervention on teacher behaviour
- Consequence for professional development in FA in the future

- Additional relevant references

What the teacher DOES, DOES NOT, and what the student DOES

	Codes for what the teacher does	Codes for student
Feedup: articulation / communication of expectancies and success criteria	1. Phase 1, general 2. Phase 1, explicate learning goals 3. Phase 1, explicate success criteria 4. Phase 1, develop learning progressions 5. Phase 1, explicitly involving students in goal / successcriteria clarification or understanding 6. NOT phase 1, teacher	1. Phase 1S, general 2. Phase 1S, actively understanding the goals/criteria 3. Phase 1S, setting /discussing own goals 4. NOT phase 1, student
Feedback: elicitation and collection of information (assessment instrument)	1. Phase 2, general 2. Phase 2, using assessment instrument to elicit student responses 3. Phase 2, consciously designing assessment methods that directly tap onto the goals 4. NOT phase 2, teacher	1. Phase 2S, develop assessment response (in general??) 2. NOT phase 2, student
Feedback: interpretation of information / judgment	1. Phase 3, general 2. Phase 3, take time to analyse student responses 3. Phase 3, reflect on student responses 4. Phase 3, relate student responses to goals 5. Phase 3, identify gaps between current performance and goals a. weaknesses b. talents 6. Phase 3, identify gaps for individual students 7. Phase 3, identify gaps at the class level 8. Phase 3, stimulate students to evaluation own/each others responses 9. NOT phase 3, teacher	1. Phase 3, general 2. Phase 3S, analysing own responses 3. Phase 3S, analysing own responses in relation to goals 4. phase 3S, analysing peer responses 5. NOT phase 3, student
Feedforward: provide feedback, communicate with students	1. Phase 4, general 2. Phase 4, provide feedback to individual students (independent of content/quality) 3. Phase 4, provide feedback that directs further improvement 4. Phase 4, stimulate students to self-reflect or peer feedback 5. NOT phase 4, teacher	1. Phase 4S, general 2. Phase 4S, provide feedback to self 3. Phase 4S, provide feedback to peers 4. Phase 4S, actively engage with feedback to understand it 5. NOT Phase 4, student
Feedup: taking action, adapt instruction, regulation of learning	1. Phase 5, general 2. Phase 5, classwide adaptation of instruction based on weaknesses 3. Phase 5, provide differentiated further instruction based on	1. Phase 5S, general 2. Phase 5S, set actions for improvement 3. Phase 5S, change learning

	analysis 4. Phase 5, classwide adaptation of instruction based on misconceptions 5. Phase 5, adapt instruction but not based on analysis 6. Phase 5, stimulate students to regulate own learning 7. Phase 6, and change (learning) behaviour 8. NOT phase 5, teacher	behaviour 4. NOT phase 5, student
relationship teacher behaviour --> effect on student	Teacher behaviour --> effect on student	
Competencies	Teacher competencies needed for FA	

Appendix 2: overzicht analyse van code “teacher behaviour → effect on student learning” (onderzoeksvraag 2)

Referentie	Docentgedrag	Fase waar dit bij hoort	Effect op student
Aschbacher, 2006	docent die geeft gemiddelde sturing door duidelijke doelen en succescriteria, vervolgens doorvragen op eerste reactie van individuele student waarbij student zelf eigen begrip moet expliciteren. Docent identificeert begrip van individuele studenten	fase 1, 2, 3 Cyclus 2-3-2-3	- hoger MC post test score (kennis) - betere science notebook scores
Ruiz-Primo & Furtak, 2006 Ruiz-Prima & Furtak, 2007	Voeren van assessment conversations volgens de ESRU cyclus (elicit student response – Student response –Recognise student response – Use information to adapt teaching) Docenten wiens assessment conversations meer voldeden aan deze cyclus hadden studenten met hoogste testscore. De meest succesvolle docent: (1) was gericht op dieper begrip van studenten, (2) stelde conceptuele vragen die conceptueel denken van studenten opriep, (3) voerde de meeste klasdiscussies om student begrip uit te diepen, (4) liet de studenten hierbij het meest zelf sturing geven aan de discussies waar zij als docent vervolgens op voortbouwde/doorvroeg, en (5) gebruikte de grootste variatie aan verzamelde informatie van studenten om vervolgvactiteiten te ontwikkelen	2-3-5 meerdere cycli van 2-3-2-3-2-3 → voldoende informatie – → naar fase 5	hogere post-test score op schriftelijke performance tests in science
Herman et al., 2015	uit de studie komt een factor “intensity of assessment use” met daarin verschillende onderdelen die allemaal met elkaar samenhangen. Hoe meer docenten hier gebruik van maken, hoe hoger het effect op de studenten. Het gaat dan om: • dagelijks lesdoelen koppelen aan curriculum doelen • Meer geplande assessments • meer en gevarieerdere methodieken gebruiken om student reacties op te wekken (aantekeningenboek, observaties, werkboekopdrachten) • meer analyseren van student werk • meer tijd nemen om te analyseren • feedback geven aan individuele studenten op basis van analyse van studentwerk • feedback aan hele klas geven • inhoud herhalen (reteaching)	onderdelen uit 1-2-3-4-5 Met focus op 3-5: analyse van werkboeken, aantekeningen boeken en observaties en op basis daarvan benoemen van een goede volgende stap	hogere post-test score op kennistoets
Conway, 2011	Docent geeft heldere doelstellingen gelijk voor alle studenten, maar met genoeg ruimte voor eigen invulling. Studenten plannen en bepalen zelf hoe en wanneer ze deze doelstellingen gaan bewijzen	fase 1 en 2	studenten nemen meer verantwoordelijkheid voor vormgeven van eigen leertraject (plannen, hoe gaan wij werken aan de leerdoelen, welke

Referentie	Docentgedrag	Fase waar dit bij hoort	Effect op student
	Stoplicht methode waarbij studenten vragen beantwoorden en aangeven dmv groen-oranje-rood in hoeverre ze zelf denken dat ze de vraag goed hebben --> studenten helpen vervolgens elkaar om samen tot een totaal plaatjes van alle vragen en doelen te komen einde van iedere les wordt in de klas 10 minuten geevalueerd: “wat hebben we nu geleerd” en het gezamenlijke leren wordt op een grote flap in de klas expliciet genoteerd	2-3-5 1-3	producten gaan we maken) en vertonen meer creativiteit in het aantonen van hun kunnen en kennen.
Willis, 2011	Docent stelt, naast inhoudelijke doelen, ook doelen op gericht op participatie en ontwerpt leeromgevingen waarin expert handelen expliciet wordt gemaakt en zet gestructureerde activiteiten in waarin studenten uitgedaagd worden om actief deel te nemen en hun rol expliciet te bediscussiëren docent zet sterk in op gelijke docent-student relatie Voorbeeld strategie: - C3B4me (see three peer students before asking me) - sage and scribe: in duo's hardop denken en aantekeningen maken	fase 1 en 2	ontwikkelen van autonomie en een autonome identiteit als lerende Student is “lid van de community”
Wingate, 2010	docenten geven gedetailleerd schriftelijke feedback (“Future gap-altering feedback”) + een indicatie van cijferrange als dit becijferd zou worden + 10 minuten gesprek daarna	fase 4	verbetering van schrijfproduct voor studenten die (1) hooggemotiveerd zijn en (2) hun eigen schrijfvaardigheid goed/hoog inschatten Dit beïnvloedt of studenten gerichte aandacht aan de feedback besteden en daarmee vervolgens hun product verbeteren
Poskitt, 2015	In Assess to Learning project ontwikkelen docenten een formatieve assessment praktijk die groeit van “niet/niet bewust toepassen van formatief toetsen” naar “bewust in praktijk brengen van formatief toetsen op 6 aspecten (zie rubric Poskitt, 2015): 1. met studenten benoemen wat er geleerd moet worden en waarom 2. met studenten benoemen hoe leren bereikt moet worden 3. met studenten onderzoeken hoe het leren eruit zou kunnen zien 4. met studenten onderzoeken hoe goed ze de leerdoelen bereikt hebben 5. klasdiscussies en student discussies over leren, toetsen en voortgang initiëren feedback, prompts en vragen stellen inzetten om leren te ondersteunen	veel gericht op fase 1, 2 en 4. Met veel aandacht voor verhelderen van leerdoelen/succescriteria, prompten en goede questionning met studenten en studenten onderling en feedback gericht op verbetering	verbeterde score op nationale writing test (nieuw zeeland)van pre naar post test ten opzichte van nationale gemiddelde + studenten die in observaties tonen meer zicht te hebben op hun eigen leren (nb kleine N)

Referentie	Docentgedrag	Fase waar dit bij hoort	Effect op student
Pham & Renshaw, 2015	Studenten geven een presentatie over een rechtenonderwerp en medestudenten moeten na afloop kritische vragen stellen. Docent stimuleert actieve participatie door aantal + kwaliteit van de vragen te scoren presentatie wordt summatief beoordeeld en formatief gebruikt	fase 2 (mondelinge presentatie + stimuleren van doorvragen door peer studenten)	studenten krijgen beter en dieper begrip van de inhoud van het vak afgemeten aan toename in complexe vragen die studenten elkaar over de stof stellen gedurende de les + studenten voelen zich zekerder in het openlijk stellen van vragen, een vaardigheid die zij cruciaal achten voor hun toekomst (method: interview)
Pham & Renshaw, 2015	Docente zet summatieve MC test formatief in: studenten moeten in groepjes elkaars antwoorden vergelijken en bij verschillen aan elkaar uitleggen. Daarna kiezen ze als hele groep het beste alternatief. Naderhand lichten de groepen hun gekozen antwoorden toe in de hele klas en docent	fase 2 -3 -2 -3	hogere score op MC kennis test over geschiedenis + beter scoren op de hogere orde denk vragen in deze toets
Pham & Renshaw, 2015	docent evalueert gedurende het vak of de gekozen werkvormen effectief zijn om de leerdoelen te bereiken in dit voorbeeld: docent ontdekt dat de studenten te weinig mogelijkheid hebben om hun vaardigheden te oefenen. Daarom vult zij de ene tussentijdse presentatie aan met een "vragen stellen strategie" waarin medestudenten verplicht na iedere presentatie kritische vragen aan elkaar moeten stellen.	fase 5	ontwikkelen van communicatie en kritische redeneervaardigheden gemeten in soort, aantal en complexiteit van de vragen die studenten elkaar stellen
Brookhart et al	Docenten experimenteren meer met een variatie aan assessment methodieken, zowel meer formeel als informeel, doen meer aan note- taking (gedurende reguliere lessen aantekeningen maken over leerlingen, oftewel: FA werd een meer geïntegreerd onderdeel van instructie) + Docenten durven af te wijken van strakke curriculum script en durven student data te gebruiken om andere strategieën toe te passen om in effectiever naar de leerdoelen toe te helpen + Vergroting van student betrokkenheid + Vieren van sterke punten van de In	Mn fase 2-4-5	Leestest scores stijgen significant meer dan in controle groep Docenten ervaren dat hun studenten meer autonoom en meer gemotiveerd worden en meer in staat zijn tot zelf monitoring Motivatatie en autonomie effect
Fluckiger et al., 2010	Casus: studenten meer betrekken bij het FT proces Casus 1	fase 2 en 4	84 ipv 37 % van de studenten rapporteert dat ze nav de quiz gaan leren wat ze nog

Referentie	Docentgedrag	Fase waar dit bij hoort	Effect op student
	three color quiz: Blad met vragen, student schrijft in zwart wat hij zelf weet, in groep vullen ze dit aan met groen, daarna nog blauw voor informatie uit tekstboeken → zicht op product, proces en voortgang en mate waarin student zelf kennis heeft, als groep of boeken nodig heeft		niet zelf blijken te weten (--> hoeveelheid blauwe inkt daalt sterk) ervaren effecten door de docent die deze casus beschrijft: meer en betere discussies in de klas, positief kritischere vragen aan elkaar, positiever klasklimaat, en de eindproducten zijn beter tov eerdere jaren
McDowell et al., 2011	vergelijking tussen meer of minder Assessment For Learning leeromgevingen. Waarbij meer AfL gekenmerkt wordt door: 1. rijkere formele feedback 2. meer informele feedback door dialoog en peer interactie 3. veel mogelijkheid om te oefenen met kennis (mn begrip) en vaardigheden 4. authentieke en relevante toetstaken 5. aandacht voor het ontwikkelen van zelfstandigheid en autonomie 6. goede balans tussen formatief en summatief toetsen	mn fase 2 en 4	Dieper leren (itt oppervlakkig leren) Studenten die in een AfL omgeving de volgende evaluaties rapporteren: meer peer support, meer feedback en ondersteuning door docenten, betere congruentie tussen doelen-leren-toetsen, meer aandacht voor gemotiveerde en diepgaander begrip van de stof Measured via student questionnaire data
Panadero et al., 2012	gebruik van rubrics (iit scripts) waarin de prestatiecriteria (leerdoelen + succescriteria) helder beschreven staan, aangevuld met process feedback. Process feedback = beschrijvende en uitleggende feedback. Niet "iets is zo" of "dat moet zo", maar "dat kun je zo doen, daar en daarom"	mn fase 1, 2 en 4	self-efficacy voor de betreffende taak (landschapsanalyse) stijgt NB. Zelf-regulatie en leren niet! Terwijl deze op vele manieren gemeten worden
Pat-el et al., 2012	Docent vertoont scaffolding gedragingen gemeten via de S-AfL-Q questionnaire. Dit zijn gedragingen die vooral gekoppeld zijn aan het geven van "how to get there"(vervolgstappen zetten) informatie.	Fase 4 en 5	Toename van intrinsieke motivatie, via eerst een positief effect op: (1) gevoel van competentie (2) gevoel van verbondenheid (Self-determination theory)
Penuel et al., 2007	Docenten die breed en frequent gebruik maken van een digitaal student response systeem. Zij gebruiken dit systeem om de volgende acties te doen: Begrip van studenten te checken (fase 2-3), diagnostische vragen stellen (fase 2), student antwoorden gezamenlijk projecteren, delen en samen bespreken (cyclus 2-3-5), feedback uit het systeem gebruiken om vervolginstructie aan te passen (fase	fase 2-3-5	Docenten ervaren effect op studenten in termen van: meer gemotiveerde studenten en daardoor een plezierige klassfeer

Referentie	Docentgedrag	Fase waar dit bij hoort	Effect op student
	5).		
Yin et al., 2008	Experimentele en controlegroep waarbij FA implementeren: geven beide zelfde science curr en exp groep implementeert “embedded assessments” zoals beschreven in Ayala et al (2008). Studie is vooral gericht op uitkomsten: bij post-test scoort exp groep hoger op motivatie, maar control group scoort hoger op achievement (mc-test en perf ass in science) Maar ... grote verschillen tussen de deelnemende docenten, niet verklaard door teaching experience en school info (b.v. laag scorende school). Niet uitgebreid geanalyseerd in deze studie, maar verschillen tussen docenten lijken te worden veroorzaakt door: - good classroom management: students can concentrate on learning - succesful teaching strategies - effective FA implementation De verklaring voor “vreemde” resultaten wordt vooral gezocht in het niet effectief implementeren van de beoogde FA strategies. Concreet hoe de FA eruit ziet => beschreven in Ayala et al (2008) in zelfde special issue.	Algemeen gericht op implementatie, niet heel specifiek de verschillende fasen onderscheiden	NIET GEVONDEN / TEGENSTRIJDIG: control groep scoort hoger op achievement. WEL: exp groep scoort hoger op motivatie
Kearney, 2013	Implementeren 2 modellen: AASL = auth ass for sustainable learning + ASPAL = auth self and peer ass for learning. Idee van het assessment process, doe docent het uitvoert is: - survey: van studentverwachtingen + engagement voor de start van de course - gezamenlijk ontwikkelen van leerdoelen voor de course, volgens backward mapping procedure. - Pilot marking: docent en studenten beoordelen gezamenlijk een aantal pieces of student work die lijken op wat studenten moeten gaan maken. Bedoeling is dat studenten criteria gaan begrijpen en dat consensus ontstaat tussen studenten en docent over te verwachten niveau en hoe je dat ziet in de verschillende pieces of student work - Peer assessment + self-asesessment. Studenten krijgen feedback van elkaar en de uitkomsten van dit proces worden met de docent besproken in een reflection session - Telt ook mee voor final grade! Als studenten meedoen met ontwikkelen van criteria, die criteria gebruiken om werk van peers (gezamenlijk) te beoordelen van feedback te voorzien, en vervolgens op basis van al die geziene werkstukken hun eigen werk beoordelen => dan zijn beter in staat om ...	Fasen 1-2-3-4	- Studenten checken beter hun eigen werk voordat ze het inleveren - Studenten kunnen de kwaliteit van hun werk beter inschatten, omdat ze vergelijkingsmateriaal van anderen hebben gezien en beoordeeld en de criteria beter begrijpen - Ability to learn autonomously => belangrijk voor sustainable learning, één van de uitgangspunten van artikel - Studenten beter reflecteren op de quality of learning => belangrijk voor sustainable learning
Lee, 2011	Docent gebruikt taakspecifieke feedback-formulieren die de aandacht van de	Fase 3 (studenten doen	- Beter understanding van eigen

Referentie	Docentgedrag	Fase waar dit bij hoort	Effect op student
	studenten sturen naar bepaalde onderdelen van de writing task (dus de focus goed sturen, gaat over writing) Studenten doen aan self & peer evaluation en noemen hierover als beneficial: - Ziet weaknesses van anderen en kunt ze helpen - Kunt leren van goede voorbeelden van anderen - Kunt je eigen schrijven verbeteren als je voorbeelden van anderen ziet Multiple drafting werkt beter dan single drafting: meer tijd om erover na te denken, kans om te verbeteren te leren van je fouten Docent stelt het geven van cijfers uit en studenten reageren hier positief op (tot verrassing van de docent):	dit) Fase 4 (feedback vd docent)	sterke en zwakke punten mbt writing - Bij zelf-evaluatie kunnen alle studenten voorbeelden nemen waar ze goed / minder goed in zijn. - Studenten gaan FA anders zien: vinden cijfers niet meer zo belangrijk, maar juist eigen sterktes / zwaktes kunnen zien en "gewoon" beter leren schrijven (een cijfers is maar een cijfer) - Studenten zien dat cijfer maar een cijfer is, en dat het ook kan leiden tot low self-esteem als vooral cijfer belangrijk vindt
James & Moreland, 2005	De docenten: - Hadden conversations met kinderen waarin ze meer expliciet en bewust stuurden op de inhoud: de gesprekken gingen meer ECHT over nature of technology, en niet alleen over "leuk een keertje technology ervaren" - In deze interactions ondersteunden de docenten ook de kinderen om bewust te zijn van HET DOEL van de activiteiten (vooraf dachten de kinderen b.v. dat het werken in een groep een belangrijker doel van de les was dan conceptual understanding van technology - Interactions waren meer gestructureerd / meer gericht doordat docenten meer in beeld hadden wat belangrijk was (intended learning goals). Daardoor konden de docenten ook beter de relatie zien tussen de leeractiviteiten en de technological goals (waarom stuur ik op deze leeractiviteit, en wat draagt dat eigenlijk bij aan het doel). - Docenten hadden meer zelfvertrouwen in FA activities, en vooral in het geval van schriftelijke feedback	Fase 1 – 2 – (3)	De kinderen waren meer geïnteresseerd, wilden in de klas graag "technological language" gebruiken, toonden meer doorzettingsvermogen, zelfvertrouwen en eigenwaarde. Kinderen lieten meer undertanding of nature of technology zien (conceptual and procedural aspects), meer technological vocabulary in de klas. Kinderen begonnen zelf knowledge gaps te zien omdat ze beter beeld hadden van learning goals + waren meer bereid om feedback te vragen.
Herppich et al., 2014	Wat doet de docent in tutor-student gesprekken waarin zij discussieren over een tekst (over human respiratory system): Gaat om verhouding tussen aantal keer dat de tutor reageert met interactive response to knowledge deficit versus geen response. - Tutoren eliciteren veel vaker een knowledge deficit, dan dat de student dat zelf spontaan expliciteert. - De tutor reageert meer / duidelijker als hij zelf de knowledge deficit heeft geëliciteerd, dan wanneer een student dit zelf/spontaan doet. - Responses van tutor zijn: (1) direct feedback = korte (neg) reactive b.v. dat iets	Fase 2 - 4	Beter begrip, gemeten in mc-test. Gesprek ging over stuk tekst over menselijk lichaam, daar ging de test ook over (dus vrij platte kennis)

Referentie	Docentgedrag	Fase waar dit bij hoort	Effect op student
	fout is, (2) scaffolding by hint or prompt, (3) give correct answer Algemeen: hoe vaker een tutor FA inzet, hoe meer de student learning gains laat zien. NB: gaat hier bij de analyses om OF de tutor reageert op een knowledge deficit, niet zozeer HOE ze dat precies doen. Als ik het goed begrijp, “tellen” alle mogelijke responses hierboven mee als een reactie van de tutor, terwijl je je af kunt vragen of b.v. correction echt een goede manier van reageren is.		
Heller et al., 2014	Drie verschillende prof dev programmes: (1) Teaching cases: analyseren het werk van studenten en gaan daarover in dialoog + analyseren (on)mogelijkheden van instructional options (allemaal op basis van tevoren geselecteerde cases) (2) Looking at student work: analyseren van het werk van EIGEN studenten + identificeren van vervolgstappen in instructie gebaseerd op gevonden bewijzen van student thinking (3) Metacognitive analysis: analyse van eigen manier van science learning en thinking (van de docent zelf dus) + identificeren van implicaties voor instructie gebaseerd op eigen ervaringen Qua science content deden allemaal hetzelfde: zelf experimenten uitvoeren om eigen conceptual understanding te verhogen, collaborative sense-making dmv discussies, illustraties van veel voorkomende misconcepties.	Fasen 2-3-5	Scores op science tests: selected response. Exp groep scoort beter dan control groep, ook jaar later en ook de English language learners. Written justification for answers: Iln moesten ook aangeven WAAROM ze een bepaald antwoord hadden gekozen. Hier waren wel verschillen tussen de prof dev programmes => Alleen “looking at student work” leidde tot verbetering bij studenten in zelfde studiejaar. Jaar later hadden (1) en (2) dit effect. Prof dev (3) had dit effect niet.
Kang et al., 2014	Docenten gebruiken 5 typen geschreven scaffolding in science tasks. In deze studie hebben ze de science assessment tasks van 33 teachers verzameld en geanalyseerd om te kijken welke scaffolds ze konden vinden. Al die scaffolds kunnen ook meer of minder sophisticated zijn: (1) Contextualised phenomena: wel of niet gecontextualiseerd met een (fictief) voorbeeld (2) Rubrics: van simpel (zoals de checklist, met punten erbij) tot uitgebreide beschrijving van niveau (3) Checklists: aantal woorden geven die voor moeten komen in antwoord / uitleg OF aantal “answer checklist” met aantal onderdeel van b.v. redenering die terug moeten komen in het antwoord. (4) Sentence frames (5) Encouraging students to draw explanatory models in combination with written explanations: varieert van vragen tekening te maken zonder verdere instructie tot modelling met een voorbeeld en een duidelijke prompt “e.g. draw a diagram ...” Het werkt ook goed om de typen scaffolding te combineren, vooral als (1) erbij zit,	Fasen 1-2	Evidence-based explanations in science tasks. De kwaliteit van die explanations nam toe bij alle 5 de typen scaffolding. Wat werkt het beste: type (1), daarna type (2), daarna type (3). Drawing (5) had negatief effect en (4) positief effect maar niet significant daarna type (5) en daarna type (4). Dus meeste effect op leren als ...

Referentie	Docentgedrag	Fase waar dit bij hoort	Effect op student
	gecombineerd met andere “higher level” scaffolding types. De kwaliteit en de combinatie van scaffolding types is belangrijker dan het aantal. Wat doen ze scaffolds: maken duidelijk wat verwachte “rigor” is, en maken daardoor taak meer “accessible” for students. Hypothese: een checklist met begrippen vermindert de cognitive load om de begrippen te vinden, en daardoor blijft er meer ruimte over voor b.v. synthetiseren / evalueren etc. Hypothese: sentence frames helpen te focussen. Studenten hebben vaak moeite de juiste taal te vinden om b.v. een redenering uit te leggen. Als er staat “things happened because ...” dan worden ze al gestuurd om een causale verklaring te geven en zijn ze meer gefocused op het daadwerkelijke probleem.		
Wiliam et al., 2004	Voor elke teacher effect size berekend, gebaseerd op voor- en nameting (die verschilt per klas/school). Docenten hebben allemaal zelf gekozen welke FA-strategie ze willen implementeren gedurende het jaar. Voorbeelden: - Questioning (door docent, In die vragen opschrijven of vragen stellen) - Feedback: b.v. alleen comments, geen cijfers - Criteria delen: doelen helder maken, targets stellen aan het begin van de les, review aan het einde van de les, voorbeelden van studentwerk - Self-assessment: stoplichten, peer-assessment, revisie in een groep - Posters / presentaties Gecategoriseerd voor: experts / moving pioniers / static pioniers / triallers => geen verschil in effect (NB: de indeling van de docenten in deze categorieën + wat de categorieën precies inhouden wordt niet duidelijk beschreven in het artikel). Algemene conclusie is hier dus: implementatie van al die FA-practices heeft in het algemeen een positief effect gehad.	Fasen 1-2 (vooral deze fasen, maar docenten konden zelf kiezen wat ze wilden implementeren / uitproberen. Geen analyse van specifieke gekozen strategie door de docent en diens effecten)	Pre-post measure + effect size: algemeen positief effect van verschillende FA-praktijken. De pre-post test verschilt per docent (konden ze zelf kiezen)
Mitton_Kukner et al., 2014	Docenten zetten writing activities in als FA-moment bij andere inhoudelijke lessen (math / science). Ze gebruiken “writing to learn” als een manier om onzichtbare proces van student thinking zichtbaar te maken. Gebruiken literacy strategies als een manier om de conversaties tussen studenten te stimuleren en hier zelf uit te leren over student progress. Voorbeelden: - Exit slips met vragen voor studenten - Mind map maken waarin connecties tussen allerlei begrippen zichtbaar moeten maken. Tijdens het maken van de mind maps loopt de docent rond en merkt op “things jumped out” dat dingen ontbreken. Gaat die dingen vervolgens bespreken en vragen of studenten dat niet begrijpen of gewoon	Fase 2-3-2-3	Studenten zijn beter voorbereid op de sum asss en scoren hier beter op

Referentie	Docentgedrag	Fase waar dit bij hoort	Effect op student
	vergeten waren.		
Hickey & Anderson, 2007	Discursive classroom ass => gaat niet om individueel ass, maar om in kaart brengen van de collectieve kennis in classroom discourse. Docent voert “feedback conversations”. Gebruiken daarbij special rubric die discourse moeten ondersteunen, gericht op “answer explanations” die redeneringen uitleggen. - Algemene voorkennis van de hele klas inventariseren - Gebruik maken van de motivatie van studenten om te weten te komen of ze een antwoord goed hebben - Answer explanations met idee van scaffolding Werken interessant idee uit van de formatieve functie van verschillende FA-practices. Verschillen tussen event-niveau, waarbij binnen paar minuten, in de discourse info volgt voor student waarop docent kan acteren => daarna dag-niveau, waarbij docent curriculum aanpast en b.v. dingen herhaalt op basis van semi-formal classroom ass. Gaat zo steeds verder in tijdsduur en van heel informeel informatie halen uit een discourse tot landelijke tests. Wat doet docent: - Activiteiten in de klas zijn duidelijk gericht op de content standards - Gebruikt close-level assessment (informal class ass, op dagniveau wat betreft tijdsspanne) en answer explanations als handvat om activiteiten vorm te geven - Gebruikt dit als handvat om op discourse-niveau (tijdsspanne van minuten, in les) goed in de gaten te houden of de gesprekken de juiste inhoud hebben - Gaat weer discourse aan nadat close-level ass zijn afgerond - Gebruiken info uit close-level en proximal assessments om te kijken naar bevestiging of aanpassing van hun idee of student understanding. Past daar weer activiteiten op aan .	Fase 2-(3)-5 (fase 3 wat minder expliciet)	Betere scores op far-transfer assessment taak
Lorente & Kirk, 2013	Wat doet docent: - Bij start van de cursus: bespreekt eerdere ass-ervaringen van studenten, vooral bij PE - Studenten discussieren in groepjes: hoe kunnen we eerder ervaren ass-problemen in deze cursus voorkomen? Groepjes van 3 a 4 studenten doen voorstellen. Enige voorwaarde vanuit de docent is dat de leerdoelen van de cursus beoordeeld moeten kunnen worden => NB: in volgende versie aangepast. Dan doet docent voorstel en mogen studenten dit bediscussiere en aanpassen. Dit kost minder tijd	Fase 1-2-3-4	Observatie van 1 docent (auteur): meer begrip bij studenten, kunnen beter onderbouwen, gebruiken FA-strategies in hun eigen klas (Teacher Educ, geobserveerd tijdens stages na dit onderzoek) Motivatie + engagement doordat meer ownership ervaren

Referentie	Docentgedrag	Fase waar dit bij hoort	Effect op student
	- Groepjes reageren op elkaars ass proposal op basis van “common criteria related to FA” ontwikkeld door de docent => doel is dat studenten van elkaar leren en hun eigen voorstellen aanpassen, zodat dit b.v. ook in hun stage kunnen gebruiken - Studenten krijgen feedback en de kans om te herzien voordat ze in hun portfolio stoppen - Student beoordeelt zichzelf, elkaar (bij team activities) en docent beoordeelt. Eventueel overleg als er teveel verschil is (shared assessment)		Meta-ass: critical thinking skills, beter inzicht in eigen sterke en zwakte punten
Gerard et al., 2010	Algemeen: met aantal Jaren prof dev neemt ook het effect op de studenten toe. Drie docenten worden gevolgd als zij 3 jaar achter elkaar in de zomer een prof dev cursus volgen en in het schooljaar het online project “plate tectonics” toepassen. Eén van de docenten is een jaar niet aanwezig bij de cursus en laat dat jaar ook minder toename zien bij haar student-scores. Wat doet docent: - Loopt meer door de klas rond om ook individueel met studenten te kunnen werken - Stopt de klas om whole class discussion te voeren als hij merkt dat veel studenten zelfde verwarring ervaren - Meer gelegenheid om te reflecteren op het eigen begrip, bijvoorbeeld door aan het einde van de dag 3 dingen te noteren die hebben geleerd - Studenten delen hun ideeën meer met klasgenoten, rapporteren aan de groep en delen hun ideeën met de docent door middel van een “exit ticket” - Analyseert het werk van studenten van de jaren ervoor. Ziet hierin dat studenten beter zijn geworden, maar dat de embedded assessment nog niet helemaal uitlokken wat hij wil weten, namelijk weer “why”-vragen. De docent past daarom de toetsen aan.	Fasen 2-3-5 en 2-3-2 (ass zelf wordt aangepast)	Delayed post-test (3 maanden na afronding). Aardrijkskunde
Buck et al., 2010	Wat deed de docent, in dit geval de docenten-in-opleiding: de meeste docenten-in-opleiding maakten in hun eigen lessen uiteindelijk geen gebruik van het FA-proces (dat aan de orde kwam in de lerarenopleiding). De succesvolle docenten lukte het om de leerlingen te “scaffolden” door discussies in kleine en grote groepen, science dagboeken, en door leerlingen tekeningen te laten maken. Hiermee monitorden ze de kennisontwikkeling van leerlingen. Ze pasten hier op ook hun onderwijs en toetsplan aan.	Fasen 2-5	Bij lerarenopleiding: sommige lio’s implementeren FA in hun eigen lessen. De leerlingen van die docenten haalden allemaal de leerdoelen.
Furtak & Ruiz-Primo, 2008	Wat doet de docent: volgens plan implementeren van “embedded ass” die bestaan uit een serie van prompts en daar tussendoor whole-class discussions. Idee is dat studenten steeds zelf aan het denken gezet worden, en dan hun ideeën delen in de whole-class discussion. Verschillende prompts:	Fasen 2-3-5	Vooraf gekeken naar hoe de verschillende prompts leiden tot meer / minder of betere whole-class discussions. De whole-class discussions zijn beter als de

Referentie	Docentgedrag	Fase waar dit bij hoort	Effect op student
	- Predict-Observe-Explain - Open vragen - Grafieken - Predict-Observe Vergelijking van de verschillende prompts gebruikt door docenten laat zien: - Open vragen laten meer en grotere variatie in misconcepties zien dan POE, waarschijnlijk omdat heel open format is. De docent krijgt dus beter beeld van de heersende misconcepties. - Open vragen geven ook de meeste input voor de whole-class discussion over de gegeven antwoorden, waarschijnlijk ook omdat de vraag open is en er daarom een diversiteit aan antwoorden zichtbaar wordt - POE leidt ook tot goede (en langere) whole-class discussions, omdat studenten gemotiveerd / enthousiast worden van experimentjes en graag hun antwoorden willen delen met de klas. - Antwoorden op schrift laten grotere variëteit (in misconcepties) zien dan de whole-class discussions. Docenten moeten dus niet alleen varen op discussies om een beeld te krijgen van student thinking.		leerlingen b.v. meer concepten gebruiken en minder misconcepties laten zien.
Phelan et al., 2012	De docenten gebruiken de POWERSOURCE assessments. Dit is een systeem van assessments én instructie-activiteiten gebaseerd op de “big ideas” van algebra. Wat doen docenten: - 1^e les: concepten uitleggen, einde van de les een toets (check for understanding) - Op basis van de uitkomsten van die toets bepalen de 2^e les, bijvoorbeeld bepaalde concepten nogmaals behandelen. Docenten konden voor deze 2^e les kiezen uit: groepswork, individueel werk of hele klas oefening (dit kozen ze uit het ontwikkelde materiaal dat ze gingen implementeren)	Fasen 1-2-5	Scores op de tussentijdse formatieve toetsmomenten (de “checks for understanding”). De experimentele groep die POWERSOURCE gebruikte scoorde hier significant beter op dan de controlegroep. Dit was zo voor alle districten die meededen, en zowel voor de short answer als de extended answer vragen in de toets
Yin et al., 2014	Hebben aantal “embedded formative assessments” ontwikkeld, gebaseerd op een learning progression. Dit houdt in: - FA helpen om het denken van leerlingen expliciet te maken, zodat ze hun eigen concepties kunnen evalueren. Dit gebeurde bij POE (predict-observe-explain) altijd eerst individueel dan delen in kleine groep, en daarna delen van groepsideeën in de hele klas in een discussie. De leerlingen moesten discussieren wat ze dachten dat er zou gebeuren. - Daarna deed de docent het experiment, zodat de leerlingen zagen wat er gebeurde (drijven of zinken) - De leerlingen schrijven individueel op wat de verschillen waren tussen hun	Fasen 2-3-5	Experimentele groep scoort beter op conceptual change, general achievement test en vooral de performance assessment. Maar niet op de mc-test. Verklaring: de nadruk bij de embedded ass lag meer op performance ass dan op kennis.

Referentie	Docentgedrag	Fase waar dit bij hoort	Effect op student
	prediction en observatie bij het experiment - Leerlingen bespreken dit in kleine groep en komen tot consensus. Dit delen ze met de hele klas, gevolgd door discussie. - Na afloop van een afgerond deel van de lessenserie krijgen de leerlingen een “challenge question” die ze triggert om de volgende conceptie in de learning progression te gebruiken (scaffolding to move to higher conceptual understanding) FA stimuleren de leerlingen om actie te ondernemen om nieuwe concepties aan te nemen, bijvoorbeeld doordat er “cognitive conflicts” waren ingebouwd in de FA die een beroep deden op verschillende concepties.		

Appendix 3. overzicht van professionaliseringstrajecten, FT-fasen en effecten op docenten (onderzoeksvraag 4)

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstraje ct expliciet op gericht is	FT-competenties	Effecten op docenten
1	Wiley & Lyon (2015)	202 wiskunde en science docenten in USA	2 jarig traject: • start workshop: werken naar eigen actieplan • workshop voor team leiders • maandelijkse meeting in Community of Practice: gezamenlijk werken aan eigen actieplannen, experimenteren met en reflecteren op FT praktijken in de klas pre / post testing	2-4-5 Namelijk via de volgende onderdelen: 1. Clarifying and sharing learning expectations 2. engineering effective classroom discussions, questions and learning tasks that elicit evidence of learning 3. providing feedback that moves learning forward 4. activating students as the owners of their own learning 5. activating students as instructional resources for one another	1. leerdoelen stellen 2. variatie aan methodieken gebruiken 3. structuur bieden aan zelf en peer beoordelingen	• toename in aantal + variatie in assessment methodieken die gebruikt worden • toename in hoeveelheid feedback • geen toename in werken met leerdoelen • kwaliteit van (1) leerdoelen (2) feedback en (3) strategieën om studenten aan te zetten tot eigen leren en (4) en peers aan te zetten als elkaars resources blijven van lage kwaliteit • meer dan 50% van de docenten doorloopt niet de hele cyclus
2	Tang et al. 2010	Case studies van docent Chinees en Engels in middelbaar onderwijs Hongkong	1 jaar • presentatie over assessment for learning en bijbehorende strategieën aan docenten en schoolleiders • 2 action learning cycles waarbij docenten experimenteren met methodieken in de klas • terugkomdagen met consultant en peer	2	inzetten van variatie aan assessment methodieken	weinig verandering in de praktijk. Docenten gebruiken de methodieken niet of niet goed wel effecten in wat docenten rapporteren geleerd te hebben over FT, echter hebben “the spirit” van FT niet geïnternaliseerd

¹ Indien in een artikel een concrete instrument opgenomen is voor het scoren/analyseren van FT gedragingen van docenten, wordt dit geel gemarkeerd in deze kolom weergegeven

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstraject expliciet op gericht is	FT-competenties	Effecten op docenten
			discussies			
3	Suurman & Koch, 2014	wiskunde docenten in Canada	twee jarig community of practice traject waarbij docenten experimenteren om de nieuwe gedachten van Assessment for Learning in Canada vorm te geven in de klas	niet gestuurd, kan vanalles zijn en wordt bepaald door wat docenten inbrengen	niet gestuurd, kan vanalles zijn en wordt bepaald door wat docenten inbrengen	experimenteren met verschillende methodieken (instrumenteel) identificeert mn dilemma's die docenten tegenkomen . Dit zijn: pedagogische, conceptuele, culturele en politieke dilemma's die allemaal aandacht behoeven wil er verandering in assessment praktijk komen
4	Furtak et al, 2008	USA science Voorgezet onderwijs 6 docenten	grootschalige implementatie van nieuw embedded assessment systematiek passend bij nieuwe science curriculum. Summier trainingsprogramma Assessment materiaal (allerlei methodieken) wordt aangeboden om te gebruiken. Dit is een volledig uitgewerkte "guide" met allerlei formatieve assessments (verschillende soorten prompts gericht op ontlocken van begrip) passend bij verschillende onderdelen en lessen van het nieuwe curriculum 6 casus studies: docenten worden intensief gevolgd om hun praktijken in kaart te brengen	2-3-5 Veel aandacht voor 2-3-2-3 cyclus met actieve groepsdiscussies (1) Eliciting student conceptions (2) Tracking and clustering student conceptions (3) Asking students to provide reasons for their explanations (4) Students argue ideas and evidence (5) Students provide evidence for their claims	- inzetten van variatie aan assessment methodieken - inzetten van groepsdiscussies - goede vragen stellen (open en conceptueel) - misconcepties identificeren op klasniveau - clusteren van studentantwoorden om trends in de klas te ontdekken	implementatie is van lage kwaliteit, met behoorlijke variatie tussen docenten: - Aangeboden assessments (prompts) worden niet gebruikt om begrip en actieve begripsdiscussies op te wekken. Mn concept maps weinig gebruikt - Prompts vaak op verkeerde moment ingezet in het curriculum - klasdiscussies als cruciaal onderdeel vaak niet ingezet. - activeren van de student gebeurt weinig
5	Rice & Smith 2011	Nieuw Zeeland 6 scholen (basis een voortgezet onderwijs) met 60 leraren	Inzetten van "Teaching as Inquiry Circle" Variatie aan activiteiten (binnen en buitenschool, individueel en in groepen, expert ondersteuning ed) om nieuwe ideeën te implementeren in de klas kwalitatieve beschrijvingen van wat	focus op fase 3 leren gebruiken van data uit de nationale taal test om sterke en zwakke punten te identificeren daarnaast doelen en		weinig veranderingen in praktijken van docenten, omdat zij sterk gericht blijven om het aanleren van inhoudelijke kennis en ze missen noodzakelijke vakdidactische kennis delen van doelen en succescriteria gebeurde nauwelijks

	Referentie + bruikbaar instrument ⁴	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstra je ct expliciet op gericht is	FT-competenties	Effecten op docenten
			docenten doen	succescriteria delen (fase 1) • effectieve feedback geven (fase 4) • goede vragen stellen (fase 2) • zelf-/ peer assessment inzetten (fase 2)		
6	Rice & Smith, 2011	daarna 1 case study van 1 docent literacy	6 weken intensief samenwerken met consultant Eerst nationale leestest scores analyseren en groepen in de klas identificeren voor iedere groep een eigen leergids ontwikkelen die studenten in hun groep doen	3 – 5	student antwoorden analyseren zwakheden en sterke punten identificeren op klasniveau voor ieder niveau passend onderwijs maken	zet gezamenlijk ontwikkelde differentiatie docent verandert de klasopstelling in groepen waardoor er direct meer student interactie komt na 6 weken significante groei op leestest voor alle groepen aan einde van jaar is effect verdwenen omdat docent niet stevig heeft doorgezet op differentiatie aanpak docent zet aanpak van analyse en differentiatie wel door, ook naar andere vakgebieden en ervaart positieve effecten op klassensfeer en attitude van leerlingen
7	Herman et al 2015	basisschool science treatment/ controle groep	context hernieuwde science curriculum met nieuwe embedded assessments met daarin tools en strategieën voor • diagnosticeren en ondersteunen van leren via observatie • analyse van student werk en zelf-	2-3-5	vakdidactische kennis geoperationaliseerd als: 1. inhoudelijke vakkennis + 2. gebruiken van geschikte en variatie aan assessment	significante groei in • vakkennis • analysevaardigheden • identificeren van vervolgstap Echter: analyses en gesuggereerde vervolgstappen waren nog van lage

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstra ject expliciet op gericht is	FT-competenties	Effecten op docenten
			reflecties eindtoetsen die summatief en ook formatief ingezet kunnen worden Professionaliseringsaanpak embedded assessment aanbod + begeleiding via 2daagse workshop ter oriëntatie op materiaal sessies om te oefenen studentmateriaal te analyseren een jaar om met materiaal te oefenen.		methodieken - kunnen analyseren van studente reacties + - kunnen kiezen van een geschikte vervolg stap voor instructie	kwaliteit Assessmentpraktijk veranderde: meer gebruik van verschillende assessmentmethodieken en hier meer tijd aan besteden ➔ alle items van “assessment use” vragenlijst vormen samen 1 factor.
8	Poskitt, 2014 Rubric, voor concrete docentprakti ken in de klas op 6 aspecten en 4 niveaus	Nieuw Zeeland basis en voortgezet onderwijs	Assessment to Learn, grootschalig (nationaal) 2 -3 jarig professionaliseringstraject gericht op verbeteren van docent kennis en assessment praktijken, maar ook aanpakken van school cultuur grote variatie aan activiteiten en werkvormen, maar de kern ligt bij Professional Learning Communities met docenten/teamleiders en externen variatie aan meetinstrumenten	1-3-4-(5) 6 doelstellingen van AtoL - met studenten benoemen wat er geleerd moet worden en waarom - met studenten benoemen hoe leren bereikt moet worden - met studenten onderzoeken hoe het leren eruit zou kunnen zien - met studenten onderzoeken hoe goed ze de leerdoelen bereikt hebben - klasdiscussies en student discussies over leren, toetsen	6 doelstellingen op 4 niveaus in Rubric beschreven, waarin groei in FA praktijk van de docent te zien is van non-awareness, awareness, internalising, applying.	Professionaliseringstraject wordt als zeer effectief beschouwd Uit observaties blijkt dat alle docenten op alle criteria groeien van non- awareness tot een van de hogere niveaus, waarbij 22% het hoogste niveau van Applying bereikt.

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstraject expliciet op gericht is	FT-competenties	Effecten op docenten
				en voortgang initiëren 11. feedback, prompts en vragen stellen inzetten om leren te ondersteunen		
9	Offerdahl & Tomanek, 2011	3 docenten hoger onderwijs, science USA	2 jarig traject waarbij 3 docenten experimenten met 2 nieuwe assessment methodieken Docenten ontwerpen met z'n drieën assessments rondom deze twee methodieken die ze alle drie uit proberen in hun lessen mix aan kwalitatieve data collectie methodieken	2 Experimenteren met Clickers en Reading Questions	• belief over wat en waarvoor assessment is • inzetten van variatie aan methodieken • student responsen vergelijken en zwakheden ontdekken • studenten betrokkener maken door nieuwe methodieken	Veranderingen in hoe docenten dachten over toetsing. • Verandering van dit zien als alleen summatief naar ook meer formatieve doelen (het verzamelen van gegevens over begrip en misvattingen van studenten) • assessment data levert input voor studenten op Maar geen verandering in hun praktijken
10	Moss et al., 2013 Observatielij st voor observeren van docentprakti jken in de klas	24 teamleiders die FT praktijken van hun docenten analyseren (486) Reading USA basis- en voortgezet onderwijs	2 jarig professioinaliseringstraject focus op vergroten van assessment kennis en leiderschap van teamleiders Teamleiders observeren de FT praktijken van hun docenten die ze zelf als “high, medium and low flyers” identificeren. Start: Film voor teamleiders over FT met focus op (1) delen leerdoelen (2) feedback geven. Teamleiders zetten dit door naar docenten, bespreken dit en gaan docenten observeren. 2-maandelijkse meetings teamleiders en onderzoekers om voortgang en observaties te bespreken en voorbeelden van FT die zij zien te	1 -4	• leerdoelen opstellen • delen van leerdoelen met studenten • geven van effectieve feedback	teamleiders: • Groeien in hun rol als “Leading Learner” en stimuleren daarmee een lerende cultuur • groeien in hun doorleeft begrip van wat FT in de klas is, hoe dit eruit ziet en hoe je daar woorden aan kunt geven • Kunnen gedetailleerder en specifiek sturen op het formuleren van leerdoelen en succescriteria in studententaal • Groeien in het geven van gedetailleerde en specifieke feedback aan docenten waarbij ze

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstraject expliciet op gericht is	FT-competenties	Effecten op docenten
			bespreken onderzoekers observeren teamleiders en geven hen ook feedback op hoe zij hun docenten ondersteunen			men focussen op leerdoelen, succescriteria en betrekken van studenten hierbij. Zijn meer gericht op wat studenten doen, naast wat docenten doen. Op docenten: teamleiders die meeste groei lieten zien, hadden scholen die meest succesvol waren in implementeren van FT in de klas (in termen van leerdoelen en betrekken van studenten hierbij)
11	Robinson et al. 2014	USA basisonderwijs verschillende vakken	2 jarige collaboratieve leergemeenschap tussen onderzoekers, assessment experts en docenten gericht op vergroten van de FT begrip, vaardigheden en praktijken van de docenten 1 jaar terugkerende bijeenkomsten of FT kennis en begrip te vergroten (dit is heel laag in begin) parallel daaraan maken docenten eigen actieplan op basis waarvan zij experimenteren in eigen praktijk en hierover discussiëren in groepmeetings kwantitatieve en kwalitatieve data over (gepercipieerde) begrip, huidige praktijken en geplande praktijken student scores op nationale kwartaal toetsen	met name 1 en 2 iets minder 3 en 4 Doelen van prof. Traject gebaseerd op FT skills van Stiggins (2006, zie Robinson et al)	- diep begrip van FT ontwikkelen - verhelderen van leerdoelen en succescriteria - gebruik van variatie aan assessment methodieken - actief betrekken van studenten bij verschillende fasen - analyseren van assessment data	ontwikkelen van diep begrip is voorwaardelijk voor - huidige gebruik van activiteiten in FT fasen - gepland gebruik van activiteiten FT fasen NB> gerapporteerd door docenten Alle items 17 items van vragenlijst over FT praktijken correleren hoog met elkaar. Docenten groeien daarmee een beetje in alle onderdelen.
12	Sabel et al., 2015 Rubrics voor	USA 49 student-docenten Science	professionalisering van inhoudelijke vakkennis (science) in combinatie met ontwikkelen van vakdidactische assesment kennis (pedagogical content knowledge)	2-3-5 lesplannen van andere docent	inzicht hebben in learning progression (hoe ontwikkelen studenten hun	inhoudelijke kennis neemt toe docenten groeien in hun vaardigheden om:

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstra je expliciet op gericht is	FT-competenties	Effecten op docenten
	scoren van docentgedra gingen gemeten via assignments mbt - Anticipating student response (F2) - Evaluating student response (F3) - Identifying next step (F5)	basisschool	Een vak waarin studentdocenten actief aan de slag gaan met didactische werkvormen en nieuwe (next generation) science curriculum in groepen van 12 en hier wekelijks een lesplan voor maken ondersteund door docenten, met vakinhoud en werkvormen MC test voor meten van inhoudelijke science kennis pre- en posttest kwantitatieve scoring van anticipating, evaluating en next step door middel van rubrics + interview achteraf	bekijken en op basis hiervan benoemen welke ideeën en misconcepties dit bij studente oproept (fase 2, “anticipating”) - begrip en misconcepties in producten van eigen studenten ontdekken (fase 3, “evaluating”) - hypothetische vervolgstap benoemen (fase 5)	kennis in dit vak) - inzicht hebben in misconcepties - misconcepties en begrip kunnen identificeren in student producten - vervolgstappen voor instructie kunnen benoemen	- te anticiperen op begrip en misconcepties - analyseren van <i>geschreven</i> studentproducten en hier misconcepties en begrip uit te halen NB. Een grafisch studentproduct kunnen ze niet goed analyseren, omdat ze hier de succescriteria niet in herkennen Ze groeien NIET op het identificeren van geschikte vervolgstap voor instructie. Dwz. De voorgestelde strategieën waren niet gepast voor de doelen die studenten moesten laten zien inhoudelijke kennis is voorspeller van FT scores
13	Panizzon & Peggs, 2008	Australie, 25 docenten nieuwe wiskunde en science curriculum voortgezet onderwijs	2 jarig professionaliseringstraject gericht op formuleren van verschillende leerdoelen door middel van SOLO taxonomie (gebaseerd op Bloom) aantal professional development sessies en continue consultatie support 2 einde van het jaar interviews met docenten kwalitatieve analyse identificeren van thema's hieruit ➔ NB gerapporteerd door docenten	1	- Ontwikkelen van doorleeft begrip van FT - formuleren van leerdoelen - zicht hebben op en kunnen ontwikkelen van learning progression (hoe ontwikkelt kennis van studenten)	Grootste winst op mate en soort van vragen stellen (fase 2) - meer gericht op verschillende niveaus van bloom en dieper begrip - meer uitvoeren van de 2-3-2-3 cyclus (vragen stellen, student reactie, doorvragen, aanvullende student reacties) belief veranderingen van summatief toetsen naar FT waarbij assessment data opleveren die vervolgonderwijs

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstrajec t expliciet op gericht is	FT-competenties	Effecten op docenten
						gericht op begrip verder kunnen sturen
14	TraughtNare & Buck, 2011	USA science voortgezet onderwijs 1 docent met 2 docentopleiders	1 jaar participatief actie onderzoek tussen 1 docent en 2 docentopleiders met 2wekelijkse reflectiesessies gedurende 2 vakken van de docent	niet gestuurd, op basis van hoe het proces gaat en de behoefte en actieplannen van de docent	belief over wat FT is	Verandering in opvatting over toetsen van cognitief proces naar relationeel proces docent gaat meer in gesprek met studenten over het stellen van vragen: studenten leren dat vragen stellen mag en dat de vragen van de docent bedoeld zijn om te helpen ipv te toetsen ➔ verandering in klasklimaat Groei in gebruik van variatie meer en betere vragen stel en discussie activiteiten die begrip van studenten prompten (fase 2) en waarbij studenten actievere rol nemen Weinig groei op: • analyseren van studentreacties, docent neemt hier tijd niet voor (NIET fase 3) • aanpassen van instructie nav studentreacties (NIET fase 5). Aanpassingen die docent doet zijn gericht op zwakheden van de klas
15	Sato et al. 2008 Rubrics voor	USA wiskunde / science voortgezet	Effecten van National Board Certification Programma. 1 jarig nationaal ontwikkeld professionaliseringsprogramme om	1-2-4-5 <i>ECHTER: inhoud wordt sterk bepaald door video's en vragen die</i>	• begrip en beliefs over wat FT is • leerdoelen formuleren	significante groei in: • begrip en belief over toetsen (nummer 1) ➔ FT meer geïntegreerd onderdeel

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsniveau)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstraject expliciet op gericht is	FT-competenties	Effecten op docenten
	scoren van docentpraktijken op 6 dimensies waarop deze studie formatief toetsen definieert	onderwijs 9 NB docenten, 7 non-NB docenten (= controlgroep)	effectiviteit van lesgeven te vergroten. groot gestructureerd programma en erkend programma (dat leidt tot certificaat) met veel materiaal, heldere doelen en verwachtingen, uitproberen in eigen praktijk, collegiale reflectiegroepen ahv video's van lessen die docenten inbrengen zowel op vakinhoud als overstijgend pre-test, na 1 jaar professionalisering, na 2 jaar (= retentie) FT gedefinieerd op 6 dimensies gemeten met rubrics + kwalitatieve data via interviews en video van lessen	<i>docenten inbrengen in hun groepen</i> 1. <i>views and uses of assessment (geen fase)</i> 2. <i>range, quality and coherence of assessment methods (fase 2)</i> 3. <i>clarity and appropriateness of goals and expectations for learning (fase 1)</i> 4. <i>opportunities for self-assessment (fasen 2-5)</i> 5. <i>modifications to teaching based on assessment information (fase 5)</i> 6. <i>quality and appropriateness of feedback (fasen 4-5)</i>	• verschillende assessment methodieken gebruiken die passend zijn voor leerdoelen (alignment) • inzetten van zelf-assessment (echter geen aandacht voor structuur hiervan) • geven van goede feedback • aanpassen van onderwijs op basis van toetsresultaten	van alle lessen • gebruik van grotere variatie aan assessment methodieken, met veel aandacht voor questionning van begrip (nummer 2, fase 2) Geen groei op • Helderheid en gepastheid van doelen en verwachtingen (nummer 3, fase 1) • kwaliteit van feedback (nummer 6, fase 4)
16	van der Pol et al., 2013	Nederland social studies, lessenserie over EU VMBO 17 docenten in experimentele groep 13 controle docenten	professionalisering via "Model of Contingent teaching" gericht op leren scaffolden als vorm van formatief toetsen 1 theoretische sessie over MCT 4 oefensessies 1 reflectieve terugkomsessie waarbij video's van lessen van de docenten gebruikt worden kwalitatieve analyse leidend tot kwantitatieve data en analyses	2-3-5 en 2-3-2-3 --> 5 cyclus 4 stappen van MCT 1. diagnose: huidige niveau (fase 2) 2. checken van begrip (fase 3) 3. scaffolding toepassen (= ondersteuning passend bij niveau (=	• gebruik van vragen stellen om begrip te ontlocken • doorvragen op begrip • vervolg (ondersteuning = scaffold) aanpassen aan niveau van student	professionalisering leidt tot groei in: • doorlopen van de 2-3-5 cyclus • ontlocken van begrip van studenten door demonstraties (fase 2) • begrip checken en doorvragen (fase 3) 2-3-2-3 cyclus • sturing aan studenten aanpassen aan niveau van de student (in

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstra je ct expliciet op gericht is	FT-competenties	Effecten op docenten
				fase 5) 4. leren checken (= terug naar fase 2)		vorm van meer of minder gestuurde vragen stellen) (fase 5) NB: minder succesvolle docenten blijken misconceptie over sturing te hebben (in FT mag je studenten niet sturen). ➔ Prof dev. Project heeft dit onvoldoende voor alle docenten kunnen aanpakken
17	Feldman & Capobianco (2008)	USA Natuur/scheiku nde voortgezet onderwijs	2 jarig collaboratief actieonderzoek gericht op het inzetten van technologie in de klas tbv formatief toetsen: • gezamenlijk kennis creëren over FT en het gebruik van technologie hierin • FT-items ontwerpen voor in het systeem • uitproberen in eigen klas • ervaringen bespreken en reflecteren • observatierapporten van onderzoekers over eigen klaspraktijken terugzien	2-3-5	• Goede items maken, passend bij leerdoelen (fase 2) • antwoorden van studenten vergelijken (fase 3) • antwoorden gebruiken om door te vragen/discussie te voeren (fase 3) • analyse van antwoorden en klasdiscussies gebruiken om instructie aan te passen	belief over wat FT is veranderde docenten veranderden van voeren van socrative dialogue (docent gestuurde vragen die zoeken naar het goede antwoord van 1 student) naar democratische klasdiscussie gekenmerkt door open vragen gericht op ontlocken en uitdiepen van begrip (fase 2 en 3 + activiveren van studenten in voeren van discussie en uitdiepen van begrip) creëren van leeromgeving waarin studenten vragen stellen en niet gericht zijn op 1 goed antwoord Nog te leren: ontwikkelen van goede, doelgerichte items die studenten uitdagen in hun zone of proximal development (NIET fase 2) analyse van antwoorden en resultaat van klasdiscussies gebruiken om

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstraject expliciet op gericht is	FT-competenties	Effecten op docenten
						vervolgstappen te zetten (NIET fase 5)
18	Mui So & Hoi Lee, 2011	Hong Kong 36 docenten science voortgezet onderwijs	research & development project om (1) kennis over gebruik van rubrics te vergoten en (2) gebruik ervan in de klas om leren te te stimuleren, te ondersteunen. professionaliseringstraject: • 3 daagse workshop vooraf, waarin docenten zelf rubrics ontwikkelen • 1 terugkom sessie tussentijds om uit te wisselen • gedurende project consultatie, observaties en feedback in de klas pre-post interventie kwalitatieve dataverzameling	2 inzetten van rubric	• ontwerpen van assesment methodiek (Rubric) • assessment inzetten om leren te stimuleren	weinig verandering in FT praktijken docenten zetten rubric alleen in voor summatieve doeleinden Rubrics hielpen om leerdoelen te communiceren naar studenten (fase 1) Rubrics niet gebruikt om feedback te geven en om vervolglernen van studenten te stimuleren (NIET fase 4 en NIET fase 5)
19	van der Pol et al. 2012	Nederland 4 docenten social studies VMBO	professionalisering via "Model of Contingent teaching" gericht op leren scaffolden als vorm van formatief toetsen 2 theoretische sessie over MCT lesobservaties met feedback en reflectieve interviews theoretische sessies over MCT en oefenen met voorbeelden kwalitatieve analyse leidend tot kwantitatieve data en analyses	2-3-5	ontlokken van begrip (diagnose, fase 2) doorvragen (fase 3) ondersteuning aanpassen aan begripsniveau van student (Fase 5)	docenten gaan meer • diagnose strategieën gebruiken, dwz open vragen stellen gericht op ontlokken van huidige begrip (fase 2) • in discussie met student ipv puur docent gestuurd vragen stellen om gezamenlijk begrip te construeren (co-construct) (fase 2) • creëren een meer open sfeer richting studenten gericht op begrip en co-constructie • ontdekken dat ze ook kunnen zoeken naar wat studenten nog niet, maar ook wat ze al wel weten (fase 3) Docenten gaan NIET meer:

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstraject expliciet op gericht is	FT-competenties	Effecten op docenten
						• begrip van studenten checken (NIET fase 3) • analyse van student discussies gebruiken om hun ondersteuning aan te passen aan niveau van de student (NIET fase 5) • de hele cyclus van MCT doorlopen (2-3-5)
20	Torrance & Pryor, 2001	UK algemeen basisonderwijs	twee cycli van actieonderzoek tussen 2 docentonderzoekers met 11 docenten. Theorie van FT bespreken en uitproberen in klas en hierop gezamenlijk reflecteren. in fase 1: gezamenlijk een model voor FT benoemen Fase 2: vanuit dit model in eigen lessen experimenteren en reflecteren variatie aan kwalitatieve dataverzameling	Niet vooraf bepaald, maar gezamenlijk gecreëerd in fase 1. Het FT model dat wordt ontwikkeld besteed aandacht aan: fase 1-2-4-5	doorleeft begrip van FT ontwikkelen	docenten veranderen mn in • verhelderen en herhaaldelijk bespreken van leerdoelen en succescriteria, niet alleen aan begin van de les (fase 1) • opstellen van inhoudelijke criteria én criteria over sociale norm/hoe werk je met elkaar samen (fase 1) • toepassen van variatie aan “vragen-stel strategieën” (fase 2) • creëren van leeromgeving waarin studenten ontdekken dat vragen stellen bedoeld is om te helpen ipv te testen (fase 2) • minder direct oordelen op eerste reacties van studenten (Fase 4) • meer ruimte creëren om eerste poging te verbeteren (Fase 5)
21	Aschbacher & Alonzo, 2006	USA Basisschool Science 8 docenten met PD 17 zonder PD	2 jarig professionaliseringstraject waarin science inhoudelijke vakkennis gecombineerd wordt met leren gebruiken van notebooks (= oefen/aantekeningen boeken) voor formatief toetsen collaboratief project van onderzoekers, experts en docenten met een aantal workshops vooraf en tijdens traject waarin docenten oefenen met inzetten en analyseren van notebooks om hier	1-2-3-4-5 (1-2-3 meest expliciet) met aandacht voor alignmet: • leerdoelen van vak • leerdoelen per les gericht op opbouwen van		2 voorwaardelijkheden: 1. Vakkennis en leerdoelen van het vak 2. Lesgeven om begrip te stimuleren ipv om het curriculum af te dekken Docenten die beide aspecten bezitten, die geven gemiddelde “just right”

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstrajec t expliciet op gericht is	FT-competenties	Effecten op docenten
			begrip van studenten uit te destilleren	kennis (learning progression) • wat voor soort oefeningen in notebooks moeten komen om aan deze leredoelen te werken • analyseren langs leerdoelen • kiezen van vervolgstap om leerdoelen te bereiken apart aandacht voor effectieve feedback strategieën		sturing aan studenten voor werken in hun oefenboeken door: • betere doelen en succescriteria (fase 1) • stellen meer open vragen gericht op begrip (fase 2) • beter kunnen identificeren van individuele zwakheden (fase 3)
22	Brookhart et al., 2010	USA taal basisschool 6 PD docenten (remedial teachers) studenten van controle groep docenten	Professionaliseringstraject met • directe instructie • eigen leerdoelen benoemen en van daaruit experimenteren en reflecteren op eigen praktijk • gezamenlijke leergroep waar drie onderzoekers feedback op geven kwalitatieve data van docenten zelf en observaties van onderzoekers	2 en 4	begrip van wat FT is	verandering van begrip/belief over wat FT is en op welke verschillende manieren je hiermee in de klas aan de slag kunt Docenten voelen zich meer bekwaam om FT vorm te geven in hun dagelijkse lespraktijk (self-efficacy) docenten gingen meer: • FT geïntegreerd toepassen in lessen, dwz. Veel meer interactief met studenten in gesprek om hun begrip te peilen en hen meer te

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstra je ct expliciet op gericht is	FT-competenties	Effecten op docenten
						betrekken en motiveren in het leerproces ➔ Sterke verandering van student- docent verhoudingen • gericht op individuele leerlingen en hun behoefte (fase 3) • (durven) afwijken van voorgeschreven curriculum, om aan te sluiten bij behoeften van studenten (fase 3 en 5)
23	Fuchs et al., 1999	USA Basisschool rekenen 8 PD docenten 8NON-PD docenten	professionalisering van docenten gericht op wiskunde onderwijs meer gericht te laten zijn op problemen oplossen ipv feiten kennis • workshop rondom ontwerpen van performance assessment voor probleem oplossen • gezamenlijk ontwerpen van 2 performance assessments • deze uitproberen in klas • in groep discussiëren over hoe deze te analyseren en te gebruiken voor aanpassen van vervolginstructie gericht op meer probleem oplossen	2-(3)-5	• ontwerpen van performance assessment gericht op doel (probleem oplossen) • identificeren van relevante vervolgstrategie	docenten groeien in hun kennis over hoe ze performance assessments voor probleem oplossen kunnen ontwerpen (Fase 2) docenten identificeren allerlei nieuwe vervolgstrategieën om meer probleem oplossen te stimuleren, MAAR deze zijn niet gebaseerd op analyse van student reponses op de performance assessments (NIET fase 3 en NIET fase 5) Hoog presterende studenten scoren beter op eindassessment, gemiddelde en laagpresterende studenten niet. DUS performance assessments worden nog niet goed formatief ingezet omdat er geen goed gevolg aan wordt gegeven.
24	Graney & Shinn, 2005	USA basisonderwijs	Professionalisering bestaat uit : • docenten beschikking geven over	3 –(5)	• analyseren van student reacties	Geen effecten. Docenten ontwikkelen geen vaardigheden om

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstrajec t expliciet op gericht is	FT-competenties	Effecten op docenten
		taal 44 docenten verdeel over 3 groepen Groep 1: krijgt feedback rapport van individuele student groep 2: krijgt feedback rapport van hele klas groep 3: krijgt geen rapport (= controle)	voortgangsscores van studenten op landelijke voltooien (mn. grafieken, tabellen) • hieraan gekoppeld een 3 uur workshop over hoe deze gegevens te interpreteren en algemene discussies over wat je hiermee zou kunnen doen in het vervolg van de lessen		• analyseren en interpreteren van grafieken en tabellen uit student/klas rapportages • kunnen identificeren en inzetten van gepaste vervolgstap passend bij analyse van studentreacties	• de rapportages zelf te interpreteren • goede vervolgstategieën te identificeren en in te zetten
25	McNaughton et al. 2012	Nieuw Zeeland basis en voorgezet onderwijs taal	hertesten van eerder effectief bewezen professionaliseringstraject "the learning school model" 3 jarig intensief professionaliseringsprogramma 2 fasen: 1. leren om achievement scores op eerdere leestesten te analyseren langs nationale normen (fase 3) en ontwikkelen van vervolginstructie die past in eigen context (fase 5) 2. gezamenlijk analyseren van nieuwe studentdata naar aanleiding van aanpassingen	3-5		leesscores van studenten gaan omhoog, en daaruit wordt geïnterpreteerd dat docenten inderdaad op een goede manier de toetscores analyseren en succesvolle vervolgstategieën weten te implementeren. NB. Weinig informatie over of en hoe dit daadwerkelijk gebeurt.
26	Chroinin and Cosgrave_20 13	Ierland Basisschool gymonderwijs case studies van	professionaliseringsprogramma met weinig interventie • Docenten krijgen toegang tot een	1-2	• ontwikkelen van doorleefte begrip van FT • formuleren van	werken met doelen en gestructureerde, geplande assessments: • geeft het vak gym meer status in

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstraj ct expliciet op gericht is	FT-competenties	Effecten op docenten
		2 + 3 docenten van twee scholen	batterij aan nieuwe assessment tools ontwikkeld bij het nieuwe gymcurriculum. • Iedere docent plant en ontwikkelt een lessenserie mbv de leerdoelen van het nieuwe gymcurriculum en selecteert hier FT assessment tools bij die passen bij deze doelen • Docenten mogen gebruiken en uitproberen wat ze willen • Iedere docent krijgt 1 onderzoeker die hij/zij kan inschakelen voor hulp en toelichting bij de tools • Docenten reflecteren op hun eigen lessen en gebruik van tools via reflectieve journals met daarin prompts gegeven door onderzoekers MN docent en student ervaringen		Leerdoelen (WALF) en succescriteria (WILF) • assessment methodieken kiezen die passen bij leerdoelen • inzetten van variatie van assessment tools	ogen van docenten en studenten ervaren • zet docenten aan tot meer reflecteren over eigen lessen en pedagogische aanpakken • docenten een grotere variatie aan assessment tools gaan gebruiken (fase 2), maar zich realiseren dat ze niet teveel in 1 les moeten willen doen. --> bewust kiezen van 1 methode passend bij doelen van de les (fase 2) Docenten rapporteren • meer kennis over FT • belief verandering naar FT om leren te stimuleren • groei in hun gym vakkennis • groei in kennis over hoe leerlingen zich ontwikkelen in hun vak (Learning progression) (Fase 1) • gegroeide vakkennis helpt bij FT processen: ontdekken van zwakheden (fase 3) en geven van gerichte feedback (fase 4) • WALT en WILF krachtig om lessen te structureren en bewust assessment tools te kiezen (fase 1-2) • Dat ze beter grip krijgen op sterke en zwakke punten van leerlingen (fase 3) • Ervaren noodzaak van bieden van structuur aan peer assessments • vaardigheden te missen om nu

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstrajec t expliciet op gericht is	FT-competenties	Effecten op docenten
						zelf assessment tools te gaan ontwikkelen (Fase 2) • Docenten gaan specifiek en gericht feedback geven passend bij de leerdoelen (fase 4)
27	Strowbridge 2008	USA 24 docenten Wiskunde, science en engels Voortgezet onderwijs	Professionaliseringstraject waarbij docenten leren omgaan met nieuwe wiskunde curriculum materiaal, inclusief de “Feedback guides” (soort rubric) voor het geven van formatieve feedback en het gebruik hiervan voor vervolgstappen Prof dev. Aanpak bestaat uit het 7 workshops gezamenlijk gespreken van • nieuwe curriculum materialen gericht op probleem oplossen • formative feedback guides en begrip van sum/formatief assessment creëren • experimenteren met feedback guide in klas en debriefing lessen daarna • continue support en coaching voor uitproberen van de feedback guides in de klas Feedback die docenten geven	(3)-4-5	• geven van specifieke feedback met acties voor verbetering • geven van feedback gericht op doelen • vervolgstategie bepalen op basis van analyse van student reacties	docenten geven geen kwalitatief goede feedback met behulp van de feedback guides. • De feedback guides lokken correctieve feedback uit en • zijn verder te ingewikkeld om goed en betrouwbaar te gebruiken • merendeel van feedback is niet gekoppeld aan leerdoelen (NIET fase 4) meer dan 50% van de voorgestelde vervolgstategieën hebben geen relatie met studentantwoorden (NIET fase 5)
28	Beesley et al., 2009 Rubrics voor scoren van	Math, USA 67 scholen, waarbinnen alle grade 4 en 5 teaches een	PD-programma: Classroom Assessment for Student Learning (Stiggins et al., 2004). Focus op: - helder communicatie van leerdoelen	Fasen 1-2-4	• heldere leerdoelen kunnen opstellen – passend bij doelen die je nastreeft • alignment kunnen	Analyse van docenthandelen obv beoordeeld studentwerk door de docent (3 math assignments) ipv observaties in de klas. De assignments bevatten: typische huiswerkopdracht

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstrajec t expliciet op gericht is	FT-competenties	Effecten op docenten
	docentgedragingen gemeten via assignments	team vormden. Scholen zijn random verdeeld in controle en exp groep	- match tussen leerdoelen en ass criteria (alignment) - feedback waarin sterke punten worden beschreven + hoe naar leerdoelen te geraken		bereiken tussen leerdoelen en assessment • feedback met sterke punten en hoe nu verder	of opdracht in de klas, typisch project of performance as, typische quiz of opdracht voor einde van de week. De docentgedragingen werden op basis van deze assignments gescoord ahv een RUBRIC. Resultaten in dit paper nog niet compleet, maar docent scores in algemeen vrij laag op: gefocusde doelen, helderheid doelen, aligment, feedback en betrekken van studenten (onderdelen van de rubric)
29	Forbes et al., 2015 Rubrics voor scoren van docentgedragingen gemeten via assignments	26 docenten Aardrijkskunde, van 13 scholen in USA (basisschool). 6 docenten van deze 26 geselecteerd voor kwalitatieve case studie: interview + video-opname in de klas. Alle docenten vulden logboek met vragen in en stuurden voorbeelden van studentwerk en	3-jarig professionaliseringstraject (studie gaat over 1 ^e jaar). Combinatie van vakinhoud en FT-praktijken. • 7 dagen summercourse • 2-maandelijks bijeenkomst in leerteams • 2 halve dagen workshop Wat ze doen: zelf onderzoekjes uitvoeren die studenten ook doen, studentwerk analyseren, FT-praktijken uit eigen klas analyseren, misconcepties identificeren in studentwerk, implementeren van vooraf bedachte FT-strategieën Methode is Reflective Assessment (Kennedy et al., 2009) die bestaat uit 4 stappen: • Anticipate student ideas • Teach a lesson • Review student artifacts • Reflect on patterns in student thinking • Adjust instruction based on	Fasen 2-3-5	• Inhoudelijke vakkennis om studentwerk te analyseren (en curriculummateriaal op waarde te kunnen schatten en te kunnen afwijken) • Relatie zien tussen vakinhoud en FT-strategieën (dat die twee met elkaar te maken hebben, dus dat algemene strategieën wel/niet werken) • Herkennen wanneer curriculummateriaal geschikt is, en wanneer niet – gerelateerd aan de doelen	Geen relatie tussen de inhoudelijke vakkennis en de kwaliteit van de FT-praktijken van docenten (was wel de hypothese) Docenten gebruik vooral curriculummateriaal om studentwerk te evalueren – in plaats eigen kennis over aardrijkskunde. Letten vooral op “plattere” dingen zoals terminologie. In fase 5 herhalen de docenten strategieën die ze kennen, en kiezen dus geen strategieën die passen bij eerdere geïdentificeerde kennishiaten.

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstrajec t expliciet op gericht is	FT-competenties	Effecten op docenten
		eigen instructies	diagnosis of students' thinking		• Durven afwijken van curriculummateriaal • Breed repertoire aan strategieën voor follow-up	
30	Furtak & Heredia, 2014	2 groepen docenten biologie (onderwerp = natuurlijke selectie), USA Springfield groep (7-9 docenten) ontwikkelt gezamenlijk met onderzoekers de learning progression. Monroe groep (3 docenten) krijgt learning progression aangeboden.	Maandelijkse bijeenkomst van groep docenten en onderzoeker. Steeds wordt learning progression besproken en gebruikt voor ontwikkelen van FT-praktijken. Stappen in professionalisering: • Gezamenlijk antwoorden van studenten analyseren om patronen te herkennen en eigen kennis van domein te vergroten • Ontwikkelen van tools om meer/beter studentreacties te ontlocken • Oefenen met de tools op elkaar en studentreacties voorspellen • Tools zelf implementeren in eigen klas • Reflecteren op implementatie door studentwerk van eigen studenten te analyseren op patronen Kern van gehele PD-traject is een learning progression.	Fasen 1-2-3-(5)	• Vakkennis over hoe begrip van studenten ontwikkelt in eigen vakgebied (b.v. in learning progression) • Gezamenlijk learning progression ontwikkelen en zo gezamenlijk betekenis geven aan hoe kennis van studenten ontwikkelt • Relateren van de learning progression aan curriculumdoelen (landelijk) • Positief beeld hebben van studenten (growth mindset) zodat FA kunnen richten op groei • Ass tools ontwikkelen op basis van learning progression	Gezamenlijk ontwikkelen en expliciteren van learning progression werkt beter dan learning progression aangereikt krijgen (negotiation of meaning, eigenaarschap). Schoonklimaat is bepalend: Monroe groep had veel accountability druk, bijvoorbeeld of de landelijke doelen wel terugkwamen in de learning progression. Hierdoor durfden docenten curr materiaal niet los te laten, terwijl Springfield docenten gezamenlijk groter repertoire aan tools ontwikkelde.

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstraject expliciet op gericht is	FT-competenties	Effecten op docenten
					• Ordenen van studentreacties volgens de learning progression	
31	Jones & Moreland, 2005	Techniekonderwijs, basisschool. 18 docenten van 5 scholen + 2 schoolleiders	Langdurig 3-jarig PD-traject gericht op vakinhoudelijke kennis over technology en FT-praktijken. In 1^e jaar werden huidige (niet effectieve) FT-praktijken geanalyseerd. Jaar 2 en 3: Docenten en onderzoekers ontwikkelen, testen en herzien gezamenlijk FT-praktijken. Inhoud PD-traject vakinhoudelijk: - Reflecteren op eigen case-studies en die van anderen - Workshops en classroom support - Teacher agreement meetings - Planning framework met learning outcomes die de kern van technology moeten weergeven Geen duidelijk PD-traject gericht op FT-praktijken beschreven. Lijkt vooral vakinhoudelijk, en vervolgens worden wel effecten beschreven op wat docenten DOEN in de klas en wat dit weer voor effect heeft op studenten.	Voornamelijk fase 1 – met effecten op fase 2 en 5, maar dit was niet expliciet het doel	• Learning outcomes formuleren die de kern van het vak weergeven • Relevante taken ontwikkelen die aansluiten bij de learning outcomes	Collaborative culture tussen onderzoekers en docenten. Positieve resultaten van studenten stimuleert andere docenten van de school om FT-praktijken ook te proberen. Planning framework heeft invloed op: bepalen van relevante learning outcomes + wat relevante taak is (gezien de doelen)
32	Barnhart & Van Es, 2015 Rubric voor scoren van "noticing" skills van de docenten op	Lerarenopleiding, science. Vergelijking van 2 cohorten: met (n = 16) en zonder video-based course (n	Cursus Learning to Learn from Teaching in lerarenopleiding. Doelen van de cursus zijn skills in 3 domeinen van "teacher noticing": - Attention to student thinking - Interpretation of student thinking - Planning and enactment of strategies to make student thinking visible	Fasen 2-3-5	• Vanuit student thinking (wat is belangrijk) activiteiten bepalen, en niet andersom • Leggen een duidelijke koppeling	Groep met cursus laat meer FT-praktijken zien: attend, analyse, respond Goede analyse van studentreactie (fase 3) en follow-up (fase 5) vereist een goede "attending to student ideas" (fase 2). Andersom geldt dit niet:

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstrajec t expliciet op gericht is	FT-competenties	Effecten op docenten
	de niveaus Lage/mediu m/hoge kwaliteit: • Attendi ng (F2) • Interpr etation (F3) • Next strateg y (F5)	= 8)	Werkwijze: - Fase 1: theorieën over responsive teaching practice - Fase 2: analyse van student thinking en docentstrategieën in video case studies - Fase 3: zelf uitproberen, opnemen en reflecteren op eigen teaching practices		tussen de fasen, dus letten op juiste thinking, analyseren die student thinking vanuit passend kader, en hebben argumenten waarom vervolgactiviteiten daarop inspelen.	sommige docenten zijn wel goed in “attend” (fase 2), maar die analyseren niet per se goed (fase 3), en hebben niet per se goede follow-up activiteiten (fase 5) Goede docenten zijn gericht op student thinking en bepalen van daaruit hun activiteiten, reageren adequaat op student ideas gedurende de les én daarna; Minder goede docenten starten bij de activiteiten en hebben geen duidelijk analyse / onderbouwing over vervolgactiviteiten
33	Heller et al., 2012	Docenten van 39 districten, over 4 condities verdeeld.	Grootschalig experiment waarin 3 verschillende PD-trajecten worden vergeleken (met ook controlegroep). Tijd: 8 * 3 uur sessies. Opzet gebaseerd en onderbouwd dmv eerdere studies naar PD Alle PD-trajecten hadden zelfde subject knowledge (science), maar andere manier van (vak)didactische kennis (zie ook tabel X) - Teaching Cases: discussies over voorgeselecteerde cases / voorbeelden uit de klas - Looking at Student Work: analyse van werk van eigen studenten en eigen FA-praktijken - Metacognitive Analysis: reflectie op eigen leerervaringen	Fasen 2-3-5	- Inhoudelijke kennis over science bezitten - Studentwerk kunnen analyseren op de belangrijkste concepten uit science (uit bepaalde unit) - Eigen docentacties kunnen analyseren / reflecteren (gezamenlijk) vervolgacties bedenken uit geanalyseerd werk van studenten	Alle 3 PD-trajecten hebben positief effect op docenten en studenten wat betreft content knowledge. Voor “justification of answers” alleen positief effect op studenten voor teaching cases en looking at student work. Focus van artikel ligt op effecten op werk van studenten.
34	Falk, 2012	11 docenten Basisschool Science: elektrische	PD-traject: 8 sessies van 3 uur (1 x per week). Elke keer verdeling van: (1) looking at student work, en (2) een science experiment (met als doel PCK ontwikkelen, dus ook	Fasen (1)-2-3-5	Vakinhoudelijke kennis, gecombineerd met ped-dic kennis (PCK)	In gebruik van FA-praktijken blijkt duidelijk relatie tussen PCK en FA- praktijken in de klas. Docenten lieten alle aspecten van FA-cyclus zien. En

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstraject expliciet op gericht is	FT-competenties	Effecten op docenten
		circuits USA	content knowledge). FA-cyclus die centraal staat in de sessies: - Doelen articuleren - Ass taken ontwerpen - Interpretieren van studentreacties - Ontwerpen van instructie In elke sessie gingen de docenten een stap verder in de FA-cyclus en elke keer brachten de docenten werk van hun eigen studenten mee. Volgorde ongeveer: - Interpreten van denken van studenten - Analyseren van task demands - Studentidee ontlocken - Verschillende manieren om proficiency te laten zien - Assesment criteria - Rubrics - Learning gaps identificeren - Instructional next steps		• Kennis over FA (als onderdeel van PCK) • Kunnen analyseren van studentreacties, gebaseerd op de doelen • Kunnen ontwerpen van passende assessmenttaken • Instructiestrategieën kunnen kiezen die aansluiten bij analyse (fase 3)	door FA-praktijken neemt ook de PCK toe. Meeste ingezette aspecten van PCK door docent: kennis van curriculum + kennis van instructiestrategieën. Gebruiken b.v. kennis van leerdoelen om studentreacties te analyseren, en om te zorgen voor alignment. Gebruiken kennis van instructiestrategieën m.n. in fase 5, en om te analyseren hoe eerdere instructie van invloed kan zijn geweest op studentreacties die analyseren.
35	Gearhart et al., 2006	3 docenten jaar lang gevolgd tijdens hun participatie in PD-traject Science, USA	PD-traject waarin portfolio met 3 belangrijke onderdelen. Portfolio bevatte ook opdrachten / instructie en daarnaast aantal bijeenkomsten per jaar met deelnemende docenten uit district. (1) Ontwerp van ass plans: reflectie op alignment tussen leerdoelen en ass taken => ben je in staat om gewenste uit te lokken (2) Analyse van studentwerk op basis van b.v. tabellen uit PD-traject => herzien van doelen en ass criteria (3) Kritische analyse van eigen FA-praktijken	Fasen 1-2-3	• Alignment tussen leerdoelen – ass taken – analyse • Heldere leerdoelen kunnen formuleren • Repertoire van formele en informele strategieën om studentreacties te ontlocken • Bewust zijn van rijkheid in informatie te vinden in studentwerk	Gingen alignment steeds beter zien gedurende PD-traject, b.v. dat analyse van studentwerk te maken heeft met hoe specifiek hun criteria zijn, en de kwaliteit van de ass taken. Gaan betere en meer heldere leerdoelen ontwikkelen, ook b.v. die verschillenden niveaus weergeven. Worden meer bewust van informatierijkheid van studentwerk.

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstrajec t expliciet op gericht is	FT-competenties	Effecten op docenten
36	Kang et al., 2014	33 science teachers USA Middelbare school Analyseren hun scaffolding in ass taken.	Geen uitgebreid PD-traject. De deelnemers zijn in hun 1 ^e jaar als teachers nog in inductieperiode. Als onderdeel daarvan hebben ze 3 sessies collegiale analyse van student werk, begeleid door docenten van uni (teacher education). In de sessies gebruike ze een rubric om studentwerk te analyseren	Fasen 1 – 2	Goede scaffolds kunnen ontwerpen in assessment taken die het voor studenten gemakkelijker / duidelijker maken wat de bedoeling is => daardoor stuur je meer richting gewenste doelen (validiteit)	De docenten gebruiken verschillende typen scaffolding in hun ass taken, en die scaffolds variëren in goed – minder goed: (6) Contextualised phenomena: wel of niet gecontextualiseerd met een (fictief) voorbeeld (7) Rubrics: van simpel (zoals de checklist, met punten erbij) tot uitgebreide beschrijving van niveau (8) Checklists: aantal woorden geven die voor moeten komen in antwoord / uitleg OF aantal “answer checklist” met aantal onderdeel van b.v. redenering die terug moeten komen in het antwoord. (9) Sentence frames (10) Encouraging students to draw explanatory models in combination with written explanations: varieert van vragen tekening te maken zonder verdere instructie tot modelling met een voorbeeld en een duidelijke prompt “e.g. draw a diagram ...”
37	Wiliam 2004	24 docenten: 12 wiskunde en 12 science (van 6 verschillende scholen). Bewust scholen geselecteerd	Bewust gekozen voor samenwerkingstraject met klein groepje docenten om echt samen interventies te bedenken (dus niet alleen tips voor teachers of docent moet uitvoeren wat ander heeft bedacht). Interventie: 8 hele dagen bijeenkomst +	Fasen 1-2-4 maar niet erg specifiek en veelal gestuurd door de vragen en praktijken van de deelnemende docenten	- Questionning technieken kunnen toepassen - Criteria delen met studenten	Gaat vooral over effecten op studenten, waarbij 4 docenten negatieve effecten laten zien t.o.v. hun eigen controlegroep (b.v. een parallelklas of collega-docent). Wat doen docenten: ze rapporteren

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstra je ct expliciet op gericht is	FT-competenties	Effecten op docenten
		die al bezig waren met FT	interview op de school, waarbij docent werd geobserveerd door onderzoekers, ideeën werden besproken over hoe beter in de praktijk uitvoeren. Onderwerpen in sessies: - Action plan maken en bespreken - Ervaringen delen - Theorieën over leren - Leerdoelen - Data-analysis (while you wait) In 1e half jaar experimenteerden de docenten met verschillende strategieën uit de interventie (b.v. questioning, comment-only marking, leerdoelen bespreken met studenten, peer/self ass). Daarna maakten eigen actieplan, passend bij hun wensen en klas.			verschillende strategieën in hun actieplan (zie ook PD-interventie), maar dit wordt niet erg specifiek uitgewerkt. Bijvoorbeeld: - Bijna alle docenten willen questioning techniques verbeteren, b.v. door open vragen, tijd om na te denken, of bij start van de les - Helft wil comment-only marking - Delen van criteria b.v. door vraag die studenten moet kunnen beantwoorden aan eind van de les, doelen noemen, laten benoemen wat geleerd hebben aan eind van de les.
38	Goertz et al., 2009	Math Basisschool 45 docenten in 9 scholen	Niet echt opgezet als PD-traject, maar elk district had een begeleider die docenten ondersteunde bij analyse van data en bij kiezen van extra / andere instructiestrategieën. Voor “instructional support” hadden ze weinig tijd. Verder hebben ze vooral vanuit schoolleider duidelijke targets en facilitering als manier om doelen te behalen: - Docenten krijgen na afloop van CA (continuous ass) tijd om te analyseren. CA moet data opleveren die gemakkelijk zijn te vergelijken over klassen / scholen heen. - Beleid en condities: heldere / strenge verwachtingen over gebruik van data	Fase (2) - 3 - 5	• Data uit toetsen (allerlei verschillende, ook CA) kunnen analyseren op betekenis voor eigen klas, maar ook breder dan dat (b.v. vergelijking binnen en buiten de school)	Docenten gebruikten de info uit CA vooral in de context van hun eigen klas en waren bijna nooit verbaasd over de uitkomsten (bevestiging van wat ze al dachten). Docenten die echt op zoek naar zijn conceptual understanding doen dit bij alle soorten toetsen => er ligt dus een soort analyse-capaciteit aan ten grondslag. Docenten die echt op zoek zijn naar conceptual change bij hun studenten zijn ook meer geneigd om andere / gevarieerdere strategieën in te zetten (in fase 5) dan re-teach => hypothese is dat deze docenten door hun analyse

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstra ject expliciet op gericht is	FT-competenties	Effecten op docenten
						(fase 3) inzien dat hun eerdere instructiestrategieën niet aansloten bij de doelen van conceptual change.
39	Gerard et al., 2010	3 docenten gevolgd als zij 3 jaar achter elkaar PD-traject volgen. Aardrijkskunde	PD-traject start vanuit perspectief van knowledge integration: idee dat nieuwe ideeën over science instructie niet zomaar bestaande ideeën kunnen vervangen, maar dat docenten daar bewijzen voor moeten zien en alternatieve instructies moeten kunnen vergelijken afwegen. Focus: koppelen van bewijzen van studentleren aan beslissingen over vervolginstructie. Gebeurt in summer courses met groepen docenten b.v.: - Ideeën van docenten expliciteren in reflectievragen en groepsdiscussies - Nieuwe ideeën aanreiken door gezamenlijk studentwerk te bekijken - Discussies met experts en collega's - Implementeren van nieuwe ideeën in volgende schooljaar Doorlopen zo aantal keer cyclus waarin projecten / assessments steeds worden herzien, geïmplementeerd en geanalyseerd.	Fasen 2-3-5 en kleincyclisch 2-3-2-3 (toets zelf wordt aangepast bij fase 5)	• Goed beeld krijgen van ALLE studenten (b.v. ook 1-op-1, kunnen rondlopen) • Toetstaken ontwerpen die doelen echt ontlocken • Groepsdiscussies stimuleren gericht op de doelen • Organisatievormen in de klas inzetten om te differentiëren, gebaseerd op geconstateerde verschillen.	Eén van de docenten is een jaar niet aanwezig bij de summer course, en laat dat jaar ook minder effecten zien op haar student-scores. Docenten verbeterden in: - Beter monitoring van studenten, b.v. door eerder formatief te toetsen, meer 1-op-1 werken, vragen stellen - Beter toetstaken, b.v. tekeningen, modellen die beter passen bij key concepts - Beter discussies in de klas, gebaseerd op FA - Beter organisatievormen in de klas, b.v. duo's, pacing van verschillende groepjes (differentiëren)

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstra je ct expliciet op gericht is	FT-competenties	Effecten op docenten
40	Gearhart & Osmundson, 2008	23 science teachers USA	Assessment Leadership Academy: PD- programma van 18 maanden, gebaseerd op portfolio met aantal opdrachten erin. Onderdelen: - Kwaliteit evalueren van eigen toetsen en “expected student responses” bedenken => gaat deze toets opleveren wat je beoogt? - Implementeren van toetstaken in eigen klas - Interpretatie van studentwerk om instructie te bepalen. Patronen analyseren, criteria opstellen	Fasen 1-2-3-5	• Toetsen kunnen ontwerpen, passend bij de doelen • Rubrics kunnen ontwerpen en gebruiken om studentwerk te analyseren • Patronen kunnen zoeken in studentwerk (vergelijken) • Breed repertoire aan instructie- strategieën.	Assessment methodieken: gaan meer coherente toetsen gebruiken, aligned met de doelen. Gedurende traject passen de docenten hun toetstaken voortdurend aan (en soms ook hun doelen) Interpreteren van studentwerk: ontwikkelen zich van eerst gewoon scoren van studentwerk, of snelle conclusies op basis van weinig waarnemingen in de klas, naar interpretatie op basis van rubrics en analyse patronen aansluitend bij doelen (echt zoeken naar wat doelen betekenen) Instructie: ontwikkelen zich van simpele strategieën zoals re-teach en correct antwoorden geven naar strategieën kiezen die beter passen, zoals peer-ass, criteria bespreken, groepsdiscussies en differentiatie. Hebben dus breder repertoire gekregen.
41	Buck et al., 2010	Teacher education Science USA	Cursus als onderdeel van science teacher education. In het onderzoek wordt deze cursus herzien met als doel om docenten-in- opleiding beter te leren gebruik te maken van FT om instructie vorm te geven. Volgens idee van teach-as-you-preach is gehele cursus ook vormgegeven volgens principes van FT en de docent doet aan “modelling” tijdens de cursus. Docenten-in-opleiding brengen vervolgens in de praktijk in hun stage. Ze moesten informatie ontlokken (fase	Fasen 2-3-5	• Begrip van FT toetsen in het algemeen: waarom is het nuttig, wat is het doel, wat is de relatie tussen FT en instructie	Pre-post verschil in begrip van docenten-in-opleiding over FT, vooral wat betreft doelen van FT en link tussen FT en instructie. Maar: konden niet in de praktijk brengen in hun stage, en zagen de voorbeelden “modelleing” in de cursus als de enige juiste mogelijkheid (ipv 1 van de vele mogelijkheden bij FT).

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstra je expliciet op gericht is	FT-competenties	Effecten op docenten
			2), dit analyseren (fase 3) en hun instructie erop aanpassen (fase 5)			
42	Phelan et al., 2012	19 docenten exp groep + 17 docenten controlegroep Algebra Middle school (11-12 jaar) USA	PD-traject: 2 dagen summercourse met oriëntatie op POWERSOURCE software. Daarna uitvoeren lessen volgens handboeken gedurende het jaar. Lessen: steeds 2 lessen van 45 min, gevolgd door 15 min toets (check for understanding). Deze lessen waren gedetailleerd uitgewerkt (door onderzoekers) in het handboek. Follow-up meetings om lessen te bespreken en data van toetsen te analyseren.	Fasen 2-3	- Verschillende methodieken inzetten (maar waren in deze studie tot in detail bedacht door onderzoekers) - Data analyseren: rank ordenen en vergelijken met andere scholen	Studenten in exp groep scoren beter op algebra testen. Docenten voeren eigenlijk vooral uit zoals bedacht: 1 ^e les concepten uitleggen, eindigen met toets en 2 ^e les baseren op deze resultaten. Vooral re-teach als dat nodig bleek en keuze uit individuele activiteiten, groepsactiviteiten of whole class.
43	Furtak, 2012	Biologie, USA Alle docenten van sectie biologie van een school. 2 jaar gevolgd tijdens traject	Zie Figuur voor centrale idee van PD-traject, rondom learning progression die doel bepaalt en daarmee ook activiteiten eromheen. Docenten werken samen in iteratief proces van ontwikkelen, uitvoeren en herzien van FT gebaseerd op de learning progression. Maandelijks bijeenkomsten: bespreken student ideas, bespreken video's van elkaars uitvoering van de FT, analyseren van studentwerk. Onderzoekers en docenten werkten voortdurend samen, aan learning progression en de FT.	Fasen 2-3-5	- Kennis van vakinhoud, maar ook van "common student ideas" daarover. - Kunnen werken met learning progression om studentreacties inhoudelijk te analyseren - Learning progression gebruiken om lessenserie vorm te geven.	Docenten gebruiken de horizontale as van de learning progression als basis voor het vormgeven van hun lessenserie (wat komt eerst, wat daarna). De verticale as gebruiken ze om misconcepties te identificeren in studentwerk en groepsdiscussies. Maar: gebruikten learning progression ook om "foute" ideeën "eruit te rammen" in plaats van ze te bespreken, b.v. door lijsten op te hangen van No-no's. Konden learning progression wel gebruiken voor fasen 2 en 3, maar niet voor fase 5.

	Referentie + bruikbaar instrument ¹	Context (land/ onderwijsnivea u)	Typering van professionaliseringsinterventie	FT-fasen/gedragingen waar professionaliseringstraje ct expliciet op gericht is	FT-competenties	Effecten op docenten
			<pre> graph TD A[Act on inferences about what student/s know] --> B[Notice or elicit student thinking] B --> C[Make inferences about what student/s know] C --> D([Purpose or Goal]) D --> A </pre>			

Appendix 4. Posters van de vijf FT-fasen en de professionaliseringsposter

Wat DOET de docent in fase 1

1. Weet kennis over leerlijn (= learning progression) in eigen vakgebied te vertalen naar goede leerdoelen en succescriteria
2. Communiceert heldere leerdoelen op meerdere manieren en meerdere momenten tijdens lessenserie
3. Stelt doelen over participatie/leren leren/autonomie
4. Gebruikt actieve werkvormen om met studenten leerdoelen en succescriteria te benoemen en verhelderen
5. Gebruikt concrete voorbeeldproducten om succescriteria uit te distilleren
6. Formuleert leerdoelen die randvoorwaarden aangeven, maar ook ruimte voor individuele invulling laten

Welke vragen stel je jezelf en elkaar?

1. Welke doelen wil ik met mijn studenten bereiken? Alleen cognitieve doelen (begrip/inzicht)?
2. Wil ik met formatief toetsen bereiken dat studenten meer verantwoordelijkheid over hun eigen leren nemen?
3. Heb ik zicht op hoe studenten kennis ontwikkelen in mijn vakgebied en welke misconcepties ze vaak hebben?
4. Werk ik met succescriteria? Hoe betrek ik studenten hierbij?
5. Hoe krijgen mijn studenten een goed beeld van de doelen en hoe ze die kunnen bereiken?
6. Is "vrijheid binnen randvoorwaarden" van toepassing op mijn leerdoelen?

Prior to commencing the writing unit, students had been exposed to and discussed a range of cartoons and comic strips. Initially, the class brainstormed pictures of cartoon and comic, and how writers 'got their message across' to readers. As these ideas were progressively refined, notes recorded them on the class whiteboard where they served as a benchmark. They returned for the duration of the unit. The teacher also provided students with an assessment rubric, that overlapped to a large extent with the student developed criteria
Hanes & Davis, 2004, p.73

De WAAT en de WOLF
• What Are we Learning For? (leerdoelen)
• What I am Looking For? (succescriteria)
Chenkin and Gograte, 2003, 1

WAGeningen UNIVERSITY & RESEARCH Judith Gullens (WUR) & Lindbeth Baartman (HU) NRC-PRO, docentnummer 405-15-722 HUISCHOOL UTRECHT

Wat DOET de docent in fase 2

1. Gebruikt manieren om studentreacties te ontlokken passend bij leerdoelen
2. Zet een variatie aan methodieken in, zowel formeel maar vooral ook veel informeel (zie kader onderaan)
3. Stelt open vragen gericht op diep begrip in plaats van op het goede antwoord
4. Stimuleert klas/groepsdiscussies om zicht te krijgen op begrip, onbegrip en misconcepties van de klas
5. Zet studenten zelf en onderling in actie om hun begrip te expliciteren en uit te diepen (dus naar meer studentsturing)
6. Kan inspringen op wat studenten inbrengen

Welke vragen stel je jezelf en elkaar?

1. Op welke manieren doe ik aan formeel en informeel formatief toetsen (zie kader)
2. Denk ik bij het kiezen van formatieve toets methode aan de leerdoelen die ik heb?
3. Welke vragen stel ik mijn studenten om te achterhalen wat hun begrip van mijn vak is? Geven die vragen mij de informatie die ik zoek?
4. Hoe kan ik een klas/groepsdiscussie gebruiken voor formatief toetsen?
5. Hoe daag ik studenten uit om hun eigen begrip te expliciteren en uit te diepen met elkaar?

Our observation of how the teacher used textbooks revealed four typical patterns of use, somewhat reminiscent of the textbook story—see table 1. 1. Informal guidance to students in what to write, which is often very vague and unsystematically focused. 2. 'Where's your answer table?' 3. Few questions that promote a little structure but still encourage free. 4. Informal guidance, the 'just right' type of questioning and direction that promote conceptual focus, often appearing in the last part of the lesson. 5. A more prescriptive guidance that promotes student copying of 'the right' information rather than exploring their own ideas and
Auchincloss & Alonzo, 2006, p.100

Table 1. Informal guidance and informal assessment practices
Formal: Designed to provide a clear, often written, working model of the learning process. Informal: Designed to provide a clear, often written, working model of the learning process. Informal: Designed to provide a clear, often written, working model of the learning process. Informal: Designed to provide a clear, often written, working model of the learning process.

WAGeningen UNIVERSITY & RESEARCH Judith Gullens (WUR) & Lindbeth Baartman (HU) NRC-PRO, docentnummer 405-15-722 HUISCHOOL UTRECHT

Wat DOET de docent in fase 3

1. Neemt de tijd om studentreacties kritisch te bekijken
2. Analyseert studentreacties op kernaspecten en leerdoelen van het vakgebied in plaats van op oppervlakkige vormenkenmerken
3. Identificeert misconcepties, sterke punten en zwakke punten op klasniveau
4. Oordeelt niet te snel, maar vraagt door en verzamelt aanvullend bewijs
5. Laat studenten actief hun eigen en elkaars werk vergelijken en interpreteren
6. Kan technologie effectief gebruiken voor formatieve doeleinden en punten 1-5

Welke vragen stel je jezelf en elkaar?

1. Op basis waarvan analyseer ik studentreacties?
2. Is het hebben van veel vak kennis handig voor het analyseren van student antwoorden?
3. Trap ik weleens in de valkuil van (te) snel oordelen? Hoe kan dat anders?
4. Zoek ik naar:
 - Individuele zwakheden?
 - Individuele sterke punten?
 - Zwakheden op klasniveau?
 - Misconcepties?
- Organiseer ik gerichte activiteiten waarmee studenten eigen en elkaars antwoorden kunnen analyseren?

Man, Judy and Brian also grew to consider class and reading questions as a mechanism for uncovering the specific conceptual difficulties students were experiencing. Previously, they had relied on their knowledge accumulated through years of teaching about general topics that students most often found difficult. Student reading questions in particular revealed much. Four main details of student understanding of these general topics as well as difficulties with concepts presented by the instructors. Other questions, on the other hand, provided evidence that what the instructors were planning about individual student understanding in one-on-one interactions actually represented a more prominent trend within the class.
Gullens & Bredewold, 2014, p. 768

Diagnose: wat vertellen de gegevens over:
• De klas? De hele klas lijkt moeite te hebben met hetgeen te maken. De klas heeft grote moeite om te reflecteren op problemen.
• Individuele leerlingen? Bij leerling 1, 7 en 13 hebben 'geen' moeite met plannen en organiseren. Leerling 2, 4 en 5 kunnen structureel reflecteren op eigen plan.
• Mijns voorbeeldend onderwijs? Bij A heb leerlingen goed voorbeeld op samenwerken en overleggen, maar moet meer aandacht besteden aan de eigen plan.
• Mijn (ons) docentschap? Bij leerlingen vragen voor tips de PBL. Het moet in meer zelfstandigheid geven in de klas.
Gullens & Bredewold (2014)

WAGeningen UNIVERSITY & RESEARCH Judith Gullens (WUR) & Lindbeth Baartman (HU) NRC-PRO, docentnummer 405-15-722 HUISCHOOL UTRECHT

Wat DOET de docent in fase 4

1. Koppelt feedback aan leerdoelen
2. Geeft concrete suggesties voor verbetering gekoppeld aan de leerdoelen
3. Biedt ruimte om iets te doen met de feedback en verbetering te laten zien
4. Laat studenten zichzelf en elkaar feedback geven, maar biedt hier een heldere structuur voor (punt 3-4, zie kader)

Welke vragen stel je jezelf en elkaar?

1. Hoe geef ik feedback? Op welke manier maak ik hierbij gebruik van de leerdoelen?
2. Wat voor een soort verbeteringsuggesties doe ik? Hoe zinvol zijn die?
3. Op welke manier geef ik studenten de mogelijkheid om zich te verbeteren na feedback? En hoe waardeer ik dat?
4. Hoe zet ik zelf- en peer-feedback in? Op welke manier geef ik hier structuur aan? Is dit de meest zinvolle vorm?

One of the teachers described how initially giving feedback concerned her self-assessment. It was one of her primary, how we were not on the green and one of the other older teachers was there with his class and I'd say four times in the space of about thirty seconds 'great job!', 'great job!', 'great job!'. And I was kind of going, 'oh I feel in my hand was 'great job!' but there's so much more going on here than just 'great job!'... (TS, interview 2)
As the lessons progressed, with teacher's feedback gradually became more specific and targeted certain aspects of the children's learning more explicitly... structuring the observation using a checklist really focused my feedback to the children and improved the quality of feedback. I was able to give them very specific areas where they were doing well and areas that they needed to improve on. (TS, reflection 6)

Table 1. Formative Assessment Cycle
1. Task: Task 1: Survey in the English class
2. Self-assessment of the whole process
3. Conference on the Task 1
4. Remedial work on Task 1
5. Task 2: Writing Task 1
6. Self-assessment of the Task 2
7. Conference on the Task 2
8. Remedial work on Task 2
9. Self-assessment of the whole process
10. Conference on the Task 2 and 3

Bij Tabel 1. Voorbeeld van hoe zelf-assessment, docent- en peer-feedback (die tijdens 'conferenties' met elkaar vergeleken worden) en mogelijkheden voor verbetering gekoppeld zijn in de formatieve toetsing.
Bredewold (2014)

WAGeningen UNIVERSITY & RESEARCH Judith Gullens (WUR) & Lindbeth Baartman (HU) NRC-PRO, docentnummer 405-15-722 HUISCHOOL UTRECHT



Wat DOET de docent in fase 5: de moeilijkste fase!

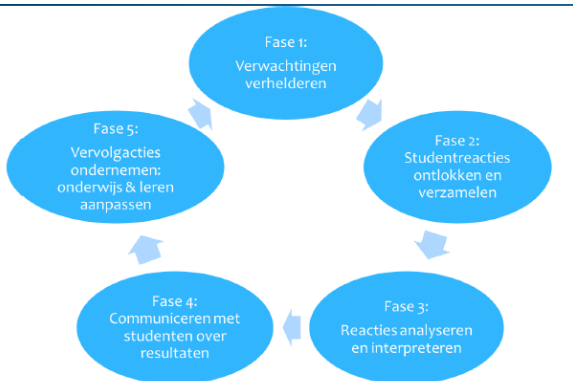
1. Past vervolginstructie/werkvorm aan naar aanleiding van analyse van studentantwoorden:
 1. Zwakten, sterktes, misconcepties
 2. Individueel, groepen, klasniveau
 3. Leerdoelen en leerlijn
2. Zet een breder didactische repertoire in dan "herhalen" en "vertragen of versnellen"
3. Durft af te wijken van voorgestructureerde curriculum/lesplan
4. Biedt structuur aan zelfregulerende/zelfsturende activiteiten

Welke vragen stel je jezelf en elkaar?

1. Pas ik mijn lessen weleens aan naar aanleiding van studentreacties?
2. Waarop baseer ik mijn keuzen dan? Analyses van studentreacties?
3. Heb ik voldoende didactisch repertoire om stof op een andere manier aan te bieden?
4. Verwacht ik dat studenten naar aanleiding van hun resultaten hun eigen leren verder sturen? Hoe help ik hen hierbij?

Wij hebben voor alle leerlingen een individueel leerplan gebaseerd op alle toetsgegevens die we van de basisschool krijgen, aangevuld met informatie van de ouders en het kind zelf. Maar natuurlijk is het onmogelijk om in de les iedere leerling individueel te begeleiden. Daarom doen wij aan "Meerkeuze": Iedere docent analyseert leerlingresultaten van de klas zijdelings op basis hiervan drie groepen te kunnen maken. Al onze docenten moeten in iedere les minimaal naar 3 niveaus differentiëren, of als in ieder geval willen leren, daarop selecteren we ook.

Locatiedirecteur VMBO-TL school



Wat behoeft nog echt professionalisering?

- Werken in de hele FT-cyclus
- Steeds zoeken naar alignment tussen de fasen
- Werken vanuit "learning progressions"
- Informeel formatief toetsen:
 - Goede vragen stellen
 - Doorvragen en klasdiscussies stimuleren met studenten aan zet
- Zwakten, sterktes en misconcepties identificeren
- FASE 5: didactisch handelingsrepertoire ontwikkelen om adequate vervolgacties te kiezen bij geïdentificeerde sterktes, zwakten en misconcepties
- De docent-student relatie expliciet aan de orde stellen

De formatieve toets-cyclus: Analysekader om te kijken naar wat de effectieve docent DOET in de klas

De kleine cyclus

Vaak gevonden in de overzichtsstudie. De effectieve docent doorloopt deze cyclus vaker, voor individuele student en hele klas, en met meer student-sturing

