

CURRICULUM EN TOETSING

“Soms zal ingeleverd moeten worden op betrouwbaarheid ten gunste van meer validiteit.”



Een curriculaire herbezinning op toetsing en examinering

Het doelgericht vormgeven van onderwijs impliceert dat tijdens en aan het eind van de rit op een passende manier wordt nagegaan in hoeverre die doelen worden gerealiseerd. Curriculum en toetsing/examinering vormen dus als het ware een tweeenheid. Maar op die principiële verbondenheid valt in de praktijk van alle dag op een aantal punten wel wat af te dingen. Wat zijn de punten waarop het knelt, wat zijn oorzaken en wat valt er vanuit curriculair perspectief aan te doen? De beantwoording van deze vragen in dit hoofdstuk mondt uit in een beargumenteerd pleidooi voor een curriculaire herbezinning op toetsing en examinering, van po tot en met de bovenbouw vo.

4. Curriculum en toetsing

Wilma Kuiper, Gerdineke van Silfhout & Bas Trimbos¹

4.1 Curriculum en toetsing: wat, waarom en waar staan we?

Verwevenheid tussen curriculum en toetsing

Curriculum en toetsing zijn principieel onlosmakelijk met elkaar verbonden. De kern van een curriculum betreft doorgaans de doelen en inhouden van leren (het 'wat' van het onderwijs), of het nu is op landelijk niveau, op school- of klasniveau of op het niveau van de individuele leerling. Die kern kan niet los worden gezien van zaken als de rol van de leraar, leeractiviteiten van leerlingen, groeperingsvormen en materialen en bronnen (het 'hoe' van het onderwijs), met een visie op leren als centrale, verbindende schakel. Het stellen van doelen impliceert dat ook wordt nagegaan en verantwoord of die worden gerealiseerd. En ook wat er aan acties geboden is als dat niet of slechts deels het geval blijkt te zijn.

SLO ontwikkelt en beproeft voorstellen voor landelijke leerplankaders en voorziet die van een wetenschappelijke onderbouwing. Samen met leraren, schoolleiders en vakexperts worden voorbeelduitwerkingen daarvan op school- en klasniveau gemaakt en uitgeprobeerd in de praktijk. Het kan daarbij gaan om exemplarische leerplanvarianten op school- en vakniveau en om voorbeeldlesmateriaal voor gebruik in de klas. Eenmaal door de overheid vastgestelde leerplankaders - kerndoelen, referentieniveaus, examenprogramma's - worden vervolgens onderhouden. In landelijke leerplankaders wordt op stelselniveau vastgelegd wat de belangrijkste doelen en inhouden zijn van een vak, leergebied of vakoverstijgend thema, sectorspecifiek of sectoroverstijgend. Daarmee wordt de grondslag gelegd voor toetsing en examinering.

Het 'wat' en het 'hoe' van het onderwijs kunnen op verschillende niveaus en vanuit verschillende perspectieven getoetst worden. Toetsing kan tot doel hebben (Sanders, 2013; ccea.org.uk):

- leraren en leerlingen informatie te verschaffen over het verloop van het leerproces bij leerlingen (formatieve toetsing ter ondersteuning van leren; hier verder aangeduid als 'formatief evalueren') en over oorzaken van eventuele stagnaties daarin (diagnostische toetsing ter ondersteuning van leren); op basis van die informatie kan vervolgens begeleiding op maat worden geboden zodat de leerling verder kan in zijn of haar leerproces.

¹ Met medewerking van Anneke Noteboom, Hans de Vries, Daniela Fasoglio, Herman Schalk, Marjo Berendsen, Jan van Hilten

- leerlingen of groepen leerlingen te beoordelen met het oog op selectie, classificering, plaatsing of certificering (summatieve toetsing, complementair aan leren). De duale examensystematiek in de bovenbouw van het vmbo en de tweede fase is hier een voorbeeld van. Dat geldt ook voor de eindtoets po, zij het dat die toets alleen adviserend van aard is.
- beleidsmakers en andere geïnteresseerden inzicht te verschaffen in de kwaliteit van het onderwijs op stelselniveau (beleidsevaluatieve toetsing), nationaal (bv. peilingsonderzoek, school op de kaart) dan wel internationaal (bv. PISA, TIMSS, PEARLS).

Verwevenheid tussen curriculum en toetsing, maar...

Als je het onderwijs ziet als een gebouw, dan zitten de (voorstellen voor) landelijke leerplankaders en de exemplarische concretisering daarvan aan de voorkant; alles wat te maken heeft met (summatieve) toetsing, examinering en verantwoording zit aan de achterkant. Dit systeem van input- en outputregulering (Kuiper, Nieveen, & Berkvens, 2013) als zodanig heeft zijn borgende waarde voor de kwaliteit van het onderwijs in de loop der jaren meer dan bewezen, alle noodzakelijk geachte tussentijdse aanpassingen daarbij inbegrepen. Toch zijn er, kijkend naar de praktijk van alledag, bij de verwevenheid van curriculum en toetsing kanttekeningen te plaatsen. Dat toetsing en examinering - in combinatie met verantwoordingsplicht richting inspectie - een sterk voorafschaduwend effect hebben op het onderwijs is een welbekend fenomeen, niet alleen in Nederland maar ook buiten onze landsgrenzen: 'getoetst maakt bemind' en 'niet getoetst maakt onbemind'. Vanwege dit terugslageffect en ook om de kwaliteit (lees: validiteit) van toetsing te bevorderen is het essentieel dat toetsing en examinering de doelen en inhouden én veranderingen daarin daadwerkelijk, eigentijds en in de volle breedte weerspiegelen (Kuiper, 2016; Kuiper et al., 2013; Leat, Livingston, & Priestley, 2013; Lundgren, 2013; Platform Onderwijs2032, 2016; Van Silfhout, 2016; Watkins, Dahlin, & Ekholm, 2005). Daarmee wordt recht gedaan aan de verbondenheid tussen curriculum en toetsing. Bovendien is het dan mogelijk de sterk sturende werking van regulering 'aan de achterkant' te benutten als hefboom voor inhoudelijke vernieuwing 'aan de voorkant' én in het onderwijsgebouw zelf. Toetsing en examinering dus *in the service of reform* (Kulm & Malcom, 1991) in plaats van als rem of blokkade. Dit belang van een toets- en examenpraktijk die meebeweegt met inhoudelijke vernieuwingen in de school- en lespraktijk wordt breed onderkend, ook op beleidsniveau (Ministerie van OCW, 2016). Tal van ervaringen in binnen- en buitenland laten evenwel zien dat dat meebewegen veel makkelijker gezegd is dan gedaan (Kuiper & Berkvens, 2013; Leat et al., 2013).

Ter toelichting hierop een tweetal voorbeelden uit eigen land. Het eerste betreft het indexamen Nederlands voor havo en vwo. Vastgelegd is dat het centraal examen alleen leesvaardigheid en argumentatieve vaardigheden toetst. Toetsing van alle

andere vaardigheden zoals kritisch en waarderend lezen, schrijfvaardigheid, literatuur, mondelinge taalvaardigheid en kennis over taal zijn toegewezen aan het schoolexamen. Het schoolexamen legt bij het bepalen van het eindcijfer evenveel gewicht in de schaal als het centraal examen. In de praktijk echter vormt het centraal examen in combinatie met het schoolexamen geen goede, evenwichtige afspiegeling van de brede doelstellingen die met het vak Nederlands worden voorgestaan. De reden daarvan is de stiefmoederlijke behandeling van de aan het schoolexamen toegewezen onderdelen. In het onderwijs gaat de aandacht vooral uit naar leesvaardigheid en argumentatieve vaardigheden, juist omdat die onderdelen centraal geëxamineerd worden. Niet centraal getoetste vaardigheden delven het onderspit. Daar komt bij dat de wijze van toetsen van leesvaardigheid in het centraal examen beperkt is: het bestaat uit het beantwoorden van grotendeels gesloten analytische vragen bij teksten.

Het tweede voorbeeld betreft de versnelde invoering (per 2018) van de verlichting van het havo-examenprogramma geschiedenis. Dat examenprogramma is gebaseerd op een voorstel van de commissie-De Rooij (uit 2001) voor een curriculum gebaseerd op oriënterende kennis van tien tijdvakken van de geschiedenis, te examineren aan de hand van een of meer casussen. Tot casustoetsing is het echter nooit gekomen. Met wat aanpassingen bleef het examen in essentie gericht op reproductie (van namen, jaartallen, begrippen en gebeurtenissen) in plaats van op historisch redeneren. Het besluit tot versnelde invoering van verlichting van het programma ontmoet stevige kritiek vanuit vakdidactische hoek. Het argument is dat de opzet van het vanaf 2018 geldende examen zodanig is dat ongeveer het omgekeerde wordt getoetst van wat met het examenprogramma wordt beoogd en daarmee het eindexamenprogramma 'onschadelijk' maakt (Wilschut, 2017).

Uitlijningsvraagstuk met diverse mogelijke oorzaken

Beide voorbeelden maken duidelijk dat er sprake is van onvoldoende afstemming of 'uitlijning' tussen curriculum en toetsing. Dergelijke uitlijningsvraagstukken kunnen verschillende (combinaties van) oorzaken hebben.

Allereerst is er het gebrek aan specificatie van doelen en inhouden. Het kan ook zijn dat doelen en inhouden überhaupt lastig te specificeren zijn, hetgeen bijvoorbeeld het geval is bij 21e eeuwse vaardigheden als kritisch denken en creatief denken. Weinig specificatie van wat leerlingen moeten kennen en kunnen en van de mate waarin dat het geval zou moeten zijn geeft veel ruimte voor eigen keuzes, maar biedt tegelijkertijd weinig houvast en richting aan toets- en examenconstructeurs, ontwikkelaars van leermiddelen en zeker ook leraren. Bovendien kan het allerlei ongewenste interpretaties in de hand werken. Gebrek aan helderheid manifesteert zich bij de door de overheid globaal geformuleerde kerndoelen voor po en onderbouw vo. Hetzelfde geldt voor de per augustus 2007

geglobaliseerde examenprogramma's voor havo en vwo. Wat dit laatste betreft is het, gezien de constructie van het centraal examen en de voorbereiding van leerlingen op dat examen, niet voor niets dat de syllabi per 2007 onverminderd gedetailleerd zijn gebleven. Er zijn twee oplossingen mogelijk voor de problemen rond het specificeren van doelen. Een door sommigen bepleite benadering is een aanpak waarbij het operationaliseren van doelstellingen en het construeren van concrete toetsopdrachten samen op gaan. Door pleitbezorgers (Millar, 2012; Wiggins & McTighe, 2006) wordt zo'n werkwijze gezien als een geschikte manier om meer zicht te krijgen op doelstellingen én beoogde leeruitkomsten. Op het punt van enkele 21e eeuwse vaardigheden is hier ervaring mee opgedaan, ook in internationaal verband. De tweede aanpak is proberen alle beslissingen over doelen en inhoud te nemen alvorens die te vertalen in toetsen en examens. In veel gevallen is dat echter onmogelijk en wellicht ook ongewenst. De meest voor de hand liggende route ligt daarom waarschijnlijk de gulden middenweg: concretiseer éérst waar mogelijk wat leerlingen moeten kennen en kunnen - ook op punten die minder makkelijk betrouwbaar en objectief meetbaar zijn, bijvoorbeeld loopbaanoriëntatie - en werk daarna uit hoe dit kan worden getoetst en verantwoord. Een dergelijke aanpak maakt zichtbaar dat toetsing het curriculum volgt, en niet omgekeerd. Zo'n werkwijze kan ook helpen de sturende werking van toetsen en examens - zoals veel scholen die nu ervaren - te verminderen (Ministerie van OCW, 2016c, p. 8).

Een tweede oorzaak kan zijn dat er sprake is van verschillen in percepties bij direct betrokkenen (vernieuwingscommissie, ontwikkelaars examenprogramma, syllabuscommissie, examenmakers, leraren) omtrent de *wenselijkheid* van inhoudelijke vernieuwingen en dus ook van het opnemen van vernieuwingsaspecten in toetsing en examinering. Dit speelde onder meer bij de ontwikkeling en beproefing van nieuwe examenprogramma's voor de natuurwetenschappelijke vakken voor de tweede fase en bij de vertaling van die programma's in pilotexamens (Kuiper, Folmer, & Ottevanger, 2012; Kuiper, Folmer, Ottevanger, & Bruning, 2011). Een vernieuwing van de examenprogramma's werd opportuun geacht om het onderwijs in deze vakken relevanter te maken voor leerlingen, meer samenhangend, minder overladen en beter aansluitend op het hoger onderwijs. Met het oog daarop werd de wisselwerking tussen contexten en concepten door de vernieuwingscommissies als gezamenlijk vertrekpunt genomen. Gaandeweg de pilots bleek echter dat daar door enkelen vraagtekens werden gezet bij het nut en de noodzaak van dit uitgangspunt dan wel bij de wijze van operationalisering ervan.

Een derde mogelijke oorzaak van onvoldoende uitlijning is dat door betrokkenen verschillend wordt aangekeken tegen de *mogelijkheid* inhoudelijke vernieuwingen valide, betrouwbaar en transparant te operationaliseren in toetsing en examinering. Dit steekt

in verschillende dossiers en met enige regelmaat de kop op. Het probleem hier is dat de validiteit van wat gemeten wordt minder gewicht krijgt, terwijl dat criterium vanuit curriculumperspectief als cruciale eis geldt. Het genoemde voorbeeld van de casustoetsing bij geschiedenis illustreert dit. Een ander voorbeeld is het onderdeel schrijfvaardigheid voor vmbo, havo en vwo: dient en kán schrijfvaardigheid getoetst worden met gebruikmaking van open, betekenisvolle schrijftaken of verdient relatief gesloten bevraging op deelaspecten de voorkeur? Belangrijk om in dit verband op te merken is dat niet alles wat relevant is voor leerlingen om te leren, op een valide, betrouwbare en transparante manier valt te toetsen (met wat voor doel dan ook) en te verantwoorden. Denk bijvoorbeeld aan zaken als loopbaanoriëntatie en -begeleiding, maatschappelijke vorming (waaronder burgerschap), het profielwerkstuk, sociaal-emotionele ontwikkeling en metacognitieve vaardigheden. Echter, ook leeropbrengsten die alleen 'merkbaar' zijn, dienen ook te worden gewaardeerd en verantwoord (Platform Onderwijs2032, 2016).

Dit uitlijningsvraagstuk en de geschetste, mogelijke oorzaken onderbouwen de relevantie van een curriculaire herbezinning op toetsing en examinering, van po tot en met de bovenbouw vo. Deze handschoen wordt door SLO opgepakt in het in 2016 gestarte project 'Toetsing en verantwoording', met de door OCW voorgenomen curriculaire herijking in po en vo als context, gelegenheid en motief. Die inhoudelijke herijking vo vraagt om "een reflectie op toetsing en examinering" (Ministerie van OCW, 2016c, p. 7). Sterker nog, een actualisatie van de onderwijsinhoud heeft pas kans van slagen als ook toetsing en examinering aangepast worden. Voorkomen moet worden dat vernieuwingen in de school- en lespraktijk worden belemmerd door onvoldoende aandacht voor een meebewegende toets- en examenpraktijk (Kuiper, 2016; Ministerie van OCW, 2016c, pp. 7/8).

4.2 Curriculaire herbezinning op toetsing en examinering

Het standpunt van SLO bij de curriculaire heroriëntatie c.q. uitdaging op het gebied van toetsing is: 'Toetsing en examinering: niet méér, maar anders en beter'. Daarmee wordt, kort aangeduid, het volgende bedoeld:

- meer aandacht voor kwaliteitsborging in termen van validiteit, betrouwbaarheid en transparantie;
- meer formatief en minder summatief evalueren;
- meer schoolgebonden toetsing en centrale toetsing wat minder prominent.

Een essentieel punt bij elk van deze drie thema's is de erkenning dat forse investeringen nodig zijn in de toerusting en ondersteuning van (toekomstige) leraren en schoolleiders. De drie thema's worden hieronder nader belicht. Per thema wordt tevens stilgestaan bij recente ontwikkelingen.

Meer aandacht voor kwaliteitsborging: validiteit, betrouwbaarheid en transparantie

Validiteit als hoeksteen

Kijkend naar de huidige praktijk valt op dat er ruim aandacht is voor (summatieve) toetsing. Een van de problemen vanuit curriculair perspectief is dat vooral getoetst wordt wat zo betrouwbaar mogelijk getoetst kan worden. Dat valt niet (altijd) volledig samen met wat vanuit de doelen getoetst zou moeten worden en welbeschouwd is dat vreemd. Validiteit is immers de hoeksteen van toetsing en examinering (Popham, 2010). Als iemand graag wil afvallen en daarom besluit een afslankdieet te volgen gecombineerd met een dagelijkse meting op een vast tijdstip van zijn of haar lichaamslengte, dan zal iedereen begrijpen dat dergelijke metingen betrouwbare en transparante maar niet-valide gegevens opleveren. Door niet of niet goed te meten wat er gemeten moet worden, verliest een toets of examen (veel van) zijn waarde. Dat wat getoetst en geëxamineerd wordt, weerspiegelt over het algemeen het 'betrouwbaar meetbare' deel van de doelstellingen. En niet alles wat van leerlingen aan kennis, vaardigheden en houdingen wordt verwacht - neem bijvoorbeeld burgerschap of sociaal-emotionele vaardigheden - is met grote betrouwbaarheid meetbaar. Maar daarmee zijn die andere, meer 'merkbare' zaken bepaald niet minder relevant.

Weerbarstige praktijk

Een positieve ontwikkeling daar waar het de versterking van de validiteit van examinering betreft is de invoering in 2016 van de beroepsgerichte profiel- en keuzevakken in het vmbo (kb, bb, gl). Profiel- en keuzevakken worden afzonderlijk van elkaar afgesloten. Profielvakken worden afgesloten met een centraal schriftelijk praktisch examen (CSPE) en mogen, indien een school daarvoor kiest, daarnaast ook met een schoolexamen worden afgesloten. Beide tellen dan voor 50% mee voor het eindcijfer. Keuzevakken worden afgesloten met alleen een schoolexamen. Het CSPE, ontwikkeld door Cito, is één examen, bestaande uit meerdere opdrachten en minitoetsen met een omvang van één of meerdere dagdelen aan het eind van schooljaar 3 of 4. Het schoolexamen bestaat net als het CSPE uit een combinatie van theoretische vragen en praktische opdrachten. Dat laatste weerspiegelt het beroepsgerichte karakter van de profielvakken en draagt daardoor bij aan meer valide examinering. Naar verluidt stemt de inhoud van het CSPE in het veld tot tevredenheid. Punten die hierbij nog aandacht behoeven zijn standaardisering van de afnamecondities en objectiviteit van de beoordeling, beide 'eigen' aan de afname en beoordeling van praktische toetsen. Andere aandachtspunten zijn: het maken van een programma van toetsing en afsluiting en kwalitatief goede schoolexamens, met name voor de keuzevakken die scholen zelf vormgeven in samenwerking en/of afstemming met mbo-instellingen, bedrijfsleven, organisaties, en

andere betrokkenen. In 2016 is een leergang schoolexaminering ontwikkeld met het doel beroepsgerichte vakdocenten op dit punt te professionaliseren. Vanuit het veld wordt overigens gevraagd om meer houvast bij het maken van schoolexamens in de vorm van in syllabi gespecificeerde eindtermen. De roep om meer houvast is begrijpelijk. De keerzijde van meer specificatie is echter dat daarmee de alom bepleite ruimte voor eigen keuzes en de schoolgebondenheid van het schoolexamen zou kunnen worden ingeperkt.

Tegelijkertijd zijn er verschillende voorbeelden te geven van discrepanties tussen wat getoetst zou moeten worden en wat feitelijk wordt getoetst. Zo ligt in zowel het leerlingvolgsysteem als de verplichte eindtoets po de nadruk tamelijk eenzijdig op rekenen-wiskunde en taal. Bij rekenen-wiskunde komen weliswaar alle domeinen en deeldomeinen terug in de toetsitems, maar de validiteit van wat gemeten wordt, ligt geregeld onder vuur. Momenteel laat het ministerie van OCW alternatieven voor de rekentoets onderzoeken, één voor het vmbo en één voor havo/vwo, met betrokkenheid van experts (wiskundeleraren en wetenschappers) aangedragen door de Nederlandse Vereniging van Wiskundeleraren (NVvW) en SLO (zie Tweede Kamer der Staten-Generaal, 2017). Bij taal ligt het accent op lezen - in casu het beantwoorden van vragen bij teksten - en op taalverzorging, en niet op productieve vaardigheden (Inspectie van het Onderwijs, 2010, 2012; zie ook de vakspecifieke trendanalyse Nederlands). Omdat zowel het leerlingvolgsysteem als de eindtoets po een belangrijke bijdrage levert aan het uiteindelijke schooladvies, is een kritische evaluatie op zijn plaats of we op deze wijze de goede zaken meten en of dat ook goed gebeurt.

In vrijwel alle door SLO uitgevoerde vakspecifieke trendanalyses (elders in deze Curriculumspiegel) wordt melding gemaakt van spanning tussen de inhoudelijke strekking van examenprogramma's en de inhoud en vorm van examens, en van spanning tussen centrale examens en schoolexamens. We noemden reeds de problemen rond het eindexamen Nederlands voor havo en vwo. Ook bij de in het vo ingevoerde rekentoets is de validiteit een stevig punt van discussie, waarbij vaak wordt genoemd dat het examen een groot beroep doet op leesvaardigheid en kennis van de wereld. Een ander voorbeeld is het meten van schrijfvaardigheid Nederlands en Engels als onderdeel van de beoogde diagnostische toetsen voor de onderbouw van het vo. Het is onder andere vanwege de adaptiviteit en een direct beschikbare diagnose dat de beheersing van deelvaardigheden getoetst worden met relatief gesloten vragen (bijvoorbeeld sleepvragen) in plaats van aan de hand van open, betekenisvolle schrijftaken. Schrijfvaardigheid in holistische zin is echter veel meer dan het beheersen van de som der deelvaardigheden. Iedereen die wel eens een tekst schrijft, zal beamen dat iemand die het 'slepen' van alinea's in de goede volgorde beheerst, niet per definitie in staat is ook een goed opgebouwde en samenhangende tekst te produceren. Het gevolg is

niet alleen een ontoereikende inhoudvaliditeit, maar ook het risico van beperkende voorafschaduwning. De beperking bestaat er dan uit dat schrijfonderwijs - voor zover daar al tijd en ruimte voor is - kan verworden tot deelvaardigheidsonderwijs dat gericht is op niet-functionele taken. Dat levert verschaald onderwijs op en een alleen maar groter wordende kloof tussen onderbouw en bovenbouw. Een soortgelijke redenering is van toepassing op gespreksvaardigheid bij de moderne vreemde talen. Zo wordt in de de vakspecifieke trendanalyse voor de moderne vreemde talen geconstateerd dat het toetsen van gespreksvaardigheid vaak (hoewel in mindere mate voor Engels) beperkt blijft tot reproductie in de vorm van min of meer uit het hoofd geleerde presentaties en dialoges. Anders gezegd, dat wat met taalverwerving wordt beoogd, wordt niet of nauwelijks getoetst.

Het probleem van valide en evenwichtige toetsing zien we op veel meer gebieden, zoals onderzoek- en ontwerpvaardigheden bij onder meer de natuurwetenschappelijke vakken, klassieke talen en klassieke culturele vorming, 21e eeuwse vaardigheden, brede vorming en beroepscompetenties. De wijze waarop leraren daarmee omgaan en de manieren van toetsen lopen sterk uiteen.

De thematiek van 'de goede dingen op een goede manier toetsen en examineren' speelt dus in alle onderwijssectoren en in de volle inhoudelijke breedte: variërend van het leerlingvolgsysteem en de eindtoets po, de diagnostische tussentijdse toetsen voor de onderbouw vo tot de examens ter afsluiting van het vmbo, havo en vwo. Het is belangrijk dat vanuit curriculumperspectief kritisch-constructief wordt meegedacht en geadviseerd over hoe in toetsing en examinering recht kan worden gedaan aan inhoudelijk relevante merkbare maar lastig(er) toetsbare onderdelen. Daar zijn ook mooie voorbeelden van te geven: het loopbaandossier in het vmbo en op het gebied van persoonlijke ontwikkeling, zoals het door TNO ontwikkelde iSelf-instrument dat inzicht geeft in hoe kinderen competenties ontwikkelen van zelfsturend leren. Tegelijkertijd laten deze en andere voorbeelden zien dat streven naar een betere balans tussen validiteit, betrouwbaarheid en transparantie betekent dat er concessies gedaan zullen moeten worden. Soms zal ingeleverd moeten worden op betrouwbaarheid ten gunste van meer validiteit, soms ook valt niet te ontkomen aan het omgekeerde.

Meer formatief en minder summatief evalueren

Formatief evalueren, een 'magic bullet'

Formatief evalueren wordt hier als synoniem gebruikt voor wat in de Angelsaksische landen wordt aangeduid als '*assessment for learning*'. We omschrijven het als: alle activiteiten die leerlingen en docent uitvoeren om de leeractiviteiten van leerlingen in kaart te brengen, te interpreteren en te gebruiken om betere beslissingen te nemen

over vervolgstappen (Black, Harrison, Lee, Marshall, & Wiliam, 2003; Black & Wiliam, 1998; CCEA, 2009). Formatief evalueren heeft als primair doel leerlingen inzicht te geven in en eigenaar te laten zijn van hun eigen leerproces en onderwijs op maat te geven (o.a. Hattie & Timperley, 2007; Sluijsmans, Joosten-ten Brinke, & Van der Vleuten, 2013; Wiliam & Leahy, 2015). Toetsing dus met leerwaarde, als integraal onderdeel van het leren (Sluijsmans et al., 2013; Sluijsmans & Kneyber, 2016) en als een proces waarbij in licht van leerdoelen de leerontwikkeling van leerlingen in kaart wordt gebracht en leerlingen feedback en begeleiding krijgen, net zolang tot ze de gewenste kennis en/of vaardigheden beheersen (zie o.a. Van Silfhout, 2016). Toetsen waarbij leren, motivatie en groei centraal staan. Andere vormen die - naast rijke feedback door de leraar, door de leerling zelf of door medeleerlingen - effectief kunnen worden ingezet zijn goede en gerichte mondelinge vragen stellen, observeren, klassengesprekken voeren, groepsdiscussies houden en reflectief lesgeven (Sluijsmans et al., 2013).

De aandacht voor formatief evalueren neemt toe op beleidsniveau en in de school- en lespraktijk van po tot en met ho. Een markante rol op dit punt is weggelegd voor de wetenschap: initiërend, expliciet de verbinding zoekend met de praktijk en het beleid, en wellicht juist daarom sterk voedend, in conceptueel en praktisch opzicht. Dat zien we internationaal (zie de vele wetenschappelijke publicaties over dit onderwerp, onder andere van Paul Black en Dylan Wiliam) en ook binnen Nederland (zie de lectoraten en publicaties van Dominique Sluijsmans, Desirée Joosten-van den Brink, Liesbeth Baartman en Kelly Meusen-Beekman). Het meest recent zijn een bundel van de hand van Sluijsmans en Kneyber (2016) en een NRO-overzichtsstudie van de hand van Gulikers en Baartman (2017). De overheid stimuleert meer formatieve evaluatie en feedback – en daarmee een betere balans tussen formatief evalueren en summatief toetsen - omdat wetenschappelijk blijkt dat het werkt. En ook omdat het scholen en docenten helpt om de ontwikkeling van de leerlingen beter in kaart te brengen en te volgen, meer maatwerk te bieden, de motivatie te verhogen van zowel leraren als leerlingen en het eigenaarschap over het leren bij leerlingen te vergroten (Ministerie van OCW, 2016b).

Deze ambitie stelt schoolleiders en docenten echter wel voor flinke uitdagingen (Hattie & Timperley, 2007). In alle sectoren van het onderwijs ligt er, zoals gemeld, veel nadruk op toetsing. Maar veel meer op summatieve dan op formatieve toetsing. Dit is niet bepaald een typisch Nederlands fenomeen (OECD, 2013). Werk maken van formatief evalueren betekent het benutten van een *'magic bullet'* (Popham, 2010): *"Assessment which is explicitly designed to promote learning is the simple, most powerful tool we have for both raising standards and empowering lifelong learners"* (Assessment Reform Group, 1999). Op dit vlak blijft tot op heden nog veel verbeterpotentieel liggen (Ministerie van OCW, 2016b).

Feedup, feedback en feedforward

Om meer zicht te krijgen op de mogelijkheden tot en condities voor inbedding van formatief evalueren in de dagelijkse school- en lespraktijk ondersteunt SLO op dit moment in het vo leraren Nederlands, Engels en wiskunde en schoolleiders van een zevental pilotscholen. Deze pilots laten tal van voorbeelden zien uit de praktijk van het vmbo, havo en vwo (Van Silfhout, 2016). Centraal in de pilots staan de vijf strategieën en bijbehorende technieken van formatief evalueren zoals uitgewerkt door Leahy, Lyon, Thomspen en Wiliam (2005; zie ook Wiliam & Leahy, 2015; zie figuur 1). Daarin draait het om drie kernvragen:

- Waar werkt de leerling naar toe, ofwel: wat zijn voor de leerling relevante leerdoelen en wat zijn de criteria om te beoordelen of leeractiviteiten succesvol zijn?
- Waar staat de leerling nu in zijn of haar leerproces ten opzichte van die leerdoelen?
- Hoe kan de leerling op basis van onder andere rijke feedback de gewenste situatie bereiken?

De eerste vraag, gericht op het verhelderen en begrijpen van leerdoelen en succescriteria/ beheersingsniveaus, heeft betrekking op wat ook wel wordt aangeduid als *'feedup'*, de tweede op het geven van terugkoppeling (*'feedback'*) en de derde op het verder helpen van leerlingen (*'feedforward'*). *Feedup* is voorwaardelijk voor *feedback* en *feedforward*. Terug naar het eerdere voorbeeld van de poging tot afslanken: de dagelijkse vinger aan de pols (*feedup*) is gebaat bij valide en betrouwbare metingen (*feedback*), maar als uiteindelijk niets met de verkregen informatie wordt gedaan (*feedforward*), schiet de exercitie aan zijn doel voorbij.

	Waar werkt de leerling naartoe?	Waar is de leerling nu?	Hoe komt de leerling naar de gewenste situatie?
Leraar	1 Verhelderen van leerdoelen en delen van criteria voor succes	2 Realiseren van effectieve discussies, taken en activiteiten die bewijs leveren voor leren	3 Feedback geven gericht op verder leren
Mede-leerling	Begrijpen en delen van leerdoelen en criteria voor succes	4 Activeren van leerlingen als belangrijke informatiebronnen voor elkaar	
Leerling	Begrijpen van leerdoelen en criteria voor succes	5 Activeren van leerlingen in het stimuleren van eigenaarschap over het eigen leren	

Figuur 1: Vijf strategieën van formatief evalueren (Leahy, Lyon, Thompson, & Wiliam, 2015).

SLO-project Toetsing en verantwoording

Het genoemde SLO-project 'Toetsing en verantwoording' beoogt meer inzicht te geven in aansprekende voorbeelden, suggesties, uitdagingen en dilemma's (ook in en uit het buitenland) rond toetsing en verantwoording. Daartoe zijn in de tweede helft van 2016 tien scholen bezocht, verdeeld over po, so, onderbouw vo, bovenbouw vmbo en tweede fase). Daarbij zijn gesprekken gevoerd met schoolleiders, leraren en leerlingen rondom thema's als valide toetsing, formatieve evaluatie, zicht op ontwikkeling, toetsbeleid en dergelijke. Tevens zijn samen met de VO-raad voor schoolleiders en leraren uit het vo twee landelijke startbijeenkomsten georganiseerd rond het thema 'Toetsen om te leren'. Die bijeenkomsten krijgen, wederom in samenwerking met de VO-raad, in 2017 een vervolg in de vorm van twee leernetwerken van in totaal ruim twintig scholen die schoolbrede invulling willen geven aan formatieve evaluatie (iets vergelijkbaars voor het po en met de PO-raad staat in de steigers). Dat alles heeft inspirerende voorbeelden naar boven gebracht van hoe formatief evalueren er in de praktijk uit ziet (SLO, 2016). Zo kan verantwoorde toetsing meer zijn dan het louter toetsen van vaststaande onderdelen waarbij het gemiddelde van verschillende eindresultaten centraal staat.

De voorbeelden uit de praktijk laten zien dat er behoefte is aan aanvulling op de (te) sterke summatieve toets- en meetcultuur in het onderwijs (zie ook Standaert, 2014). Ze demonstreren bovendien dat meer investeren in formatief evalueren kan betekenen dat minder zwaar geleund hoeft te worden op summatieve toetsing, al blijft dat laatste uiteraard ook nodig (zie verderop). Elke toetsvorm - of het nu gaat om een overhoring, proefwerk, luistertoets, presentatie, werkstuk, examen, observatie, een vorm van peer feedback of anderszins, formeel dan wel informeel van aard - kan worden beschouwd als bron van informatie. Afzonderlijk geven dergelijke bronnen te weinig informatie voor een zwaarwegende beslissing over leerlingen, bijvoorbeeld of zij kunnen overgaan of niet, kunnen op- of afstromen, of hun diploma mogen behalen. Maar samen kunnen ze de bewijslast leveren voor beslissingen over de voortgang van een leerling (Van der Vleuten, Schuwirth, Driessen, Govaerts, & Heeneman, 2015).

Investeren in een feedbackcultuur

Het enthousiasme voor een meer op feedback gerichte cultuur in het onderwijs neemt toe. Tegelijkertijd zijn er veel vragen en knelpunten (Sluijsmans & Kneyber, 2016). In twee door PLEXS (het landelijk platform voor examensecretarissen), SLO en VO-raad georganiseerde bijeenkomsten over formatieve evaluatie in oktober 2016 noemden leraren en schoolleiders verschillende aspecten. Voor leerlingen en ouders is het wennen. Van leerlingen - vooral die in het vo - wordt immers een actieve(re) rol verwacht. Ouders kijken soms vreemd op als hun kinderen minder vaak of niet meer met cijfers thuis komen, zeker als ze gewend zijn - en soms eerder dan hun kind(eren) zelf - in

Magister of een vergelijkbaar systeem te kijken welke cijfers zijn behaald. Van leraren en schoolleiders wordt ook veel gevraagd, eerst en vooral een heldere, gedeelde visie op leren en de rol van toetsing daarbinnen in combinatie met een cultuurverandering (feedbackcultuur). Maar ook is bijpassende professionalisering nodig. Op dat punt valt er veel te winnen (zie bijvoorbeeld Adie & Willis, 2016; Willis & Adie, 2016), op curriculair, didactisch en pedagogisch vlak. Gulikers en Baartman (2017) noemen de volgende competenties als voorwaardelijk: een doorleefd begrip van formatieve evaluatie, vakinhoudelijke kennis en inzicht in leerlijnen en leerontwikkeling kunnen toepassen, een vakdidactisch handelingsrepertoire breder dan herhalen en vertragen kunnen inzetten, leerling-docentrelaties expliciet aan de orde kunnen stellen met een focus op leerling-sturing en technologie effectief kunnen inzetten ter ondersteuning van het formatieve evaluatieproces. Een noodzakelijke voorwaarde voor effectieve feed-up, feedback en feedforward is daarom ook het vertalen van landelijke leerplankaders naar concrete leerlijnen en duidelijke leerdoelen en succescriteria. Maar dat is ingewikkeld en bewerkelijk. Op beleidsniveau en ook in de praktijk dient zich bovendien een lastig dilemma aan: concretisering van landelijke leerplankaders in toetsbare leerdoelen verschaft meer helderheid, richting en houvast, maar overdaad aan specificatie schaadt. De ervaringen in Engeland (Leat, Livingston, & Priestley, 2013) en Schotland (Wallace & Priestley, 2017) zijn daar het levende bewijs van. Alles dichttimmeren kan de gewenste ruimte voor eigen keuzes ernstig beknotten en blijkt bovendien vaak de opmaat tot meer summatieve toetsing (Nieveen & Kuiper, 2013). Daar komt het risico bij - zo leren ook de schoolbezoeken in het kader van het project 'Toetsing en verantwoording' - dat overzichten van leerdoelen kunnen verworden tot lesgeven volgens afvinklijstjes waardoor de samenhang tussen leerdoelen verdwijnt. Leren rekenen in het po bijvoorbeeld is meer dan het leren en aftikken van een lijst losse leerdoelen. Van belang is dat leerlingen ook oog krijgen voor de samenhang (het 'cement') tussen de verschillende leerdoelen. En die samenhang is lastig in losse leerdoelen te vangen.

Bij (het investeren in) een feedbackcultuur speelt ook reflectie een belangrijke rol. Tiener College in Gorinchem - een van de tien bezochte scholen - biedt een doorlopende leerlijn voor kinderen van 10 tot 14 jaar. De school hecht veel waarde aan reflectie. Drie keer per jaar zijn er PLP-gesprekken gepland tussen leraar, leerling en ouders. Bij elk gesprek stelt de leerling samen met leraar en ouders persoonlijke leerdoelen op in de vorm van een persoonlijk leerplan (PLP). Leerdoelen kunnen bijvoorbeeld liggen op sociaal-emotioneel en/of cognitief vlak. De uitdaging voor Tiener College is nu om deze persoonlijke doelen en de dagelijkse leeractiviteiten nog meer met elkaar te verbinden, zodat leerlingen ervaren dat alle activiteiten bijdragen aan het behalen van gestelde doelen, ze continu zicht hebben op waar ze staan en erop kunnen reflecteren. Maar ook dat ouders niet pas tijdens PLP-gesprekken echt zicht krijgen op het leren en de ontwikkeling van hun kind.

Instrumenten voor formatieve evaluatie

In opdracht van het ministerie van OCW wordt momenteel een onderzoek uitgevoerd naar vraag naar (vanuit scholen) en aanbod van formele instrumenten en interventies op het gebied van formatief evalueren. De resultaten van dit onderzoek komen in het voorjaar van 2017 beschikbaar. Een voorbeeld van een instrument in wording is 'Groeimeter'.

Dit is een bij Cito in onderzoek zijnde online platform om de groei van leerlingen in het basisonderwijs te monitoren. Het doel van Groeimeter is leerkrachten en leerlingen inzicht te geven in de leerdoelen die leerlingen dienen te beheersen en feitelijk beheersen. Dit stelt leraren en leerlingen in staat individuele leerroutes samen te stellen. Leerlingen krijgen zo meer regie over hun eigen leerproces doordat zij werken aan afgesproken leerdoelen en zelf (mee) bepalen wanneer zij klaar zijn om hun kennis en vaardigheden te demonstreren. Dit doen ze door opdrachten te maken, bijvoorbeeld een gesprek voeren, een tekening maken en online vragen beantwoorden. Na afloop kunnen zij checken of ze het juiste antwoord hebben gegeven. Ze zien dan ook direct of/dat zij weer een stap hebben gezet en wat hun volgende leerdoel wordt. Iedere leerling kan in zijn eigen tempo werken. Daarnaast bestaat de mogelijkheid dat leerlingen gelijktijdig aan bepaalde leerdoelen werken in de klas of bijvoorbeeld op een excursie. Cito gaat dit instrument beproeven met een groep basisscholen.

Een ander voorbeeld van een formatief instrument is de diagnostische tussentijdse toets (DTT) voor Engels, Nederlands en wiskunde, bedoeld voor afname aan het einde van de onderbouw vo en ontwikkeld door Cito, onder regie van CvTE. Inmiddels is de DTT overgedragen aan de markt. Dedact en A-vision & Universiteit Maastricht gaan de toets aanbieden. De toets heeft een formatief doel: informatie bieden over de individuele leerbehoeftes van leerlingen, zodat het onderwijsleerproces kan worden bijgestuurd (Ministerie van OCW, 2016b). De toets is gekoppeld aan de tussendoelen vo, onderscheidt vijf niveaus (vmbo-bb, vmbo-kb, vmbo-gt, havo en vwo), en is adaptief en digitaal. Oorspronkelijk werd ingezet op verplichte afname, maar onder druk van de Tweede Kamer is in 2014 gekozen voor een pilot waar scholen op vrijwillige basis aan mee konden doen. Tijdens deze door SLO ondersteunde pilot heeft de DTT op een aantal scholen als vliegwiel gediend om formatief evalueren een plek te geven. Figuur 2 plaatst de DTT in de context van formatief evalueren. De vijf grijs gekleurde fases om de roze kern omvatten dagelijkse informele activiteiten (*day-to-day, minute by minute*) van formatief evalueren. De DTT is gepositioneerd als een van de formele instrumenten in de buitencirkel.



Figuur 2: DTT in relatie tot formatief evalueren.

De bedoeling is dat de DTT inzicht geeft in de sterke en zwakke punten van een leerling (*feedback*) en dat mede op basis van observaties en andere informele activiteiten de docent samen met een leerling kan nagaan of de diagnose klopt en kan bepalen wat een vervolgstap kan zijn (*feedforward*). Een belangrijk aandachtspunt, zo blijkt uit de pilots, is de nog geringe aandacht van leraren voor het verhelderen van leerdoelen en succescriteria (*feedup*). Zonder deze basis lijken *feedback* en *feedforward* zinloos of relatief willekeurig. Beide zijn immers direct gerelateerd aan leerdoelen en succescriteria. De DTT is niet onomstreden. Ook uit Kamervragen (Ministerie van OCW, 2016b) blijkt dat er twijfels zijn over of de toets in de praktijk gebruikt kan en gaat worden zoals beoogd.

Meer schoolgebonden toetsing en centrale toetsing wat minder prominent

Eindtoets po

Om overgangen in het onderwijs binnen en tussen sectoren te borgen en om kansenongelijkheid tegen te gaan zijn ook vormen van objectieve centrale toetsing en examinering van belang (OECD, 2015). Die vormen kennen we in Nederland ook. Zo is er de verplichting voor alle leerlingen van groep 8 in het reguliere basisonderwijs een eindtoets te maken. Deze eindtoets meet of leerlingen de referentieniveaus voor taal en rekenen beheersen. De eindtoets is geen examen waar een leerling voor kan slagen of zakken, maar

geeft een advies 'best passend brugklatype' naast het schooladvies. De toelating van een leerling tot een vo-school mag niet afhangen van het resultaat van die leerling op een eindtoets po. Het schooladvies is leidend en kan alleen naar boven worden bijgesteld als de eindtoets daar aanleiding toe geeft. Met de invoering per 2014 van de Wet Eindtoetsing PO hebben ook marktpartijen de mogelijkheid gekregen een eindtoets aan te bieden. Dat heeft erin geresulteerd dat er inmiddels - naast de Centrale Eindtoets van CvTE/Cito - ook door andere toetsaanbieders ontwikkelde eindtoetsen beschikbaar zijn: Route 8 (een digitale, adaptieve eindtoets van A-VISION), de IEP Eindtoets (Bureau ICE), de Dia-eindtoets, de CESAN Eindtoets en de AMN Eindtoets. In elke eindtoets worden eigen accenten gelegd, zodat scholen kunnen kiezen welke toets het beste bij hun onderwijsvisie en doelgroep past. Voor wat betreft de Centrale Eindtoets kan in 2017 gekozen worden voor een papieren of een digitale versie. Vanaf 2018 wordt een beweging in gang gezet naar een digitale adaptieve Centrale Eindtoets.

Welke eindtoets door scholen ook wordt gebruikt, van belang is het objectieve karakter van zo'n toets. Schooladviezen hebben als pre dat ook niet-meetbare aspecten worden meegenomen en niet gebaseerd zijn op een momentopname. De vermindering van het gewicht van de eindtoets ten gunste van dat van het schooladvies heeft als mogelijke keerzijde dat de minder objectieve schooladviezen in het voordeel uitpakken van toch al bevoorrechte groepen leerlingen. Op dat punt is er reden tot zorg. Scholen zijn verplicht het schooladvies te heroverwegen als de eindtoets hoger uitvalt dan dit advies, maar lang niet alle scholen doen dit ook: slechts één op de vijf kinderen krijgt een aangepast advies. Uit de tussenevaluatie van de eindtoets blijkt dat het schooladvies voor groep 8-leerlingen in toenemende mate samenhangt met het opleidingsniveau van de ouders van leerlingen. Kinderen met laagopgeleide ouders krijgen structureel een lager advies dan kinderen met hoogopgeleide ouders, ook bij een vergelijkbare score op de eindtoets. Het systeem lijkt zo kinderen met laagopgeleide ouders te benadelen. Een ander punt van zorg is de onderlinge vergelijkbaarheid van de eindtoetsen (Ministerie van OCW, 2016a). Uit een analyse van toetsresultaten (Emons, Glas, & Berding-Oldersma, 2016) bleek dat de resultaten in 2016 op de drie eindtoetsen met op dat moment het grootste marktaandeel (Centrale Eindtoets, Route 8, IEP Eindtoets) uiteenliepen, met verschillen in schooladvisering als gevolg. Maatregelen als aangepaste normeringen, meer afstemming bij het opzetten en uitvoeren van pre-testactiviteiten, en het gebruik van één gezamenlijke set ankeropgaven (per 2019) hebben tot doel aan deze zorg tegemoet te komen. De vraag is evenwel in hoeverre dat alles ook wenselijk is. Immers, het toelaten van meerdere toetsen zou, zo was de gedachte, een impuls moeten geven aan verbetering van de eindtoetsing in de zin van: dingen anders meten. De voorgestelde maatregelen lijken hiermee enigszins op gespannen voet te staan.

Centraal examen en schoolexamen in bovenbouw vmbo en tweede fase

De bovenbouw van het vmbo en de tweede fase kennen een systeem van eindtoetsing in de vorm van centrale examens en schoolexamens, beide 50%. Deze duale systematiek vormt een in potentie krachtig duo, omdat het alle gelegenheid biedt voor toetsing en examinering die naar inhoud en vorm recht doet aan een breed, samenhangend curriculum. De eerlijkheid gebiedt echter te melden dat dit tweetal behoorlijk met elkaar uit de pas loopt als gevolg van overwaardering van het centrale en onderwaardering van het schoolgebonden deel.

Centrale toetsing en examinering maakt op landelijk niveau de leeropbrengsten bij vakken zichtbaar. Daarbij wordt duidelijk hoe een individuele leerling presteert ten opzichte van het gemiddelde van zijn of haar onderwijsniveau. Ook is het mogelijk de gemiddelden van scholen met elkaar te vergelijken. Bij centrale examens worden niet alleen hoge eisen gesteld aan validiteit maar ook aan betrouwbaarheid. Leerlingen die een vergelijkbare prestatie leveren moeten hetzelfde cijfer krijgen, ongeacht de vraag wie hen beoordeelt en op welk moment het examen wordt afgenomen (vergelijkbaarheid over jaren). Centrale examens zeggen dus iets over de kwaliteit van het onderwijs in termen van leeropbrengsten, op leerling-, school- en stelstelniveau.

Centrale examinering kent echter ook een aantal beperkingen. Het louter schriftelijke karakter van de centrale examens voor havo en vwo legt beperkingen op aan dat wat getoetst kan worden. In de praktijk blijken centrale examens vooral kennis en inzicht te toetsen, en in beperkte mate hogere denkvaardigheden. Vanuit een streven naar zo hoog mogelijke betrouwbaarheid gebeurt dat, afhankelijk van het vak, voor een belangrijk deel op basis van een gesloten wijze van vragen. Schoolexamens bieden meer ruimte voor open, betekenisvolle taken en activiteiten, maar die ruimte wordt vaak onvoldoende benut. Voor het vmbo geldt voor de algemeen vormende vakken hetzelfde, met dien verstande dat het reeds genoemde centraal schriftelijk praktijk examen bij de beroepsgerichte vakken naast kennisonderdelen ook een praktijkopdracht bevat. Een andere beperking is dat centrale toetsen en examens noodgedwongen slechts een deel van het geheel toetsen. Niet alle te toetsen, daartoe geormerkte eindtermen kunnen in één centraal schriftelijk afgenomen examen aan de orde komen. Beschikbare tijd en begrensde belastbaarheid van leerlingen maken dat onmogelijk.

Enige relativering van het centraal examen

Gezien deze beperkingen ligt enige relativering van centrale examinering in de rede. 'Enige relativering' betekent nadrukkelijk niet hetzelfde als 'schrappen'. Voor wat betreft het laatste en aansluitend op wat de Onderwijsraad (2015) stelt in zijn advies over 'Maatwerk binnen wettelijke kaders': meer maatwerk in het stelsel is wenselijk onder gelijktijdig handhaven van ijkpunten, waaronder degelijke en betrouwbare centrale eindtoetsing. Het centraal examen alleen kan onvoldoende zichtbaar maken wat de opbrengsten van een beoogd curriculum in de volle breedte zijn. Het centraal examen geeft een indicatie van het bereikte niveau van kennis (in de zin van: weten dat, weten hoe, weten waarom) omtrent bepaalde delen uit het examenprogramma en voor de beroepsgerichte vakken ook voor wat betreft het praktijkniveau. Tegelijkertijd zegt het weinig tot niets over de beheersing van overige kennis, vaardigheden en attitudeontwikkeling.

Herwaardering en kwaliteitsborging van schoolexamen

Schoolexamens bieden wél de mogelijkheid andere toetsvormen te hanteren. Daar zijn de schoolexamens ook voor bedoeld. Immers, bij vakken die met een centraal examen en een schoolexamen worden afgesloten, worden met het schoolexamen (ook) andere kennis en vaardigheden getoetst en beoordeeld dan met het centraal examen. Recht willen doen aan doelen en inhouden in de volle breedte vraagt om een herwaardering van het schoolexamen; en daarmee om een betere verdeling en balans tussen wat getoetst zou moeten worden in het centraal examen en in het schoolexamen en in hoe dat zou kunnen gebeuren. Het schoolexamen is de verantwoordelijkheid van de school. Die balans lijkt op dit moment zoek, wetende dat het niet ongebruikelijk is schoolexamens - waarvoor de verantwoordelijkheid bij de school ligt - naar inhoud en vorm te modelleren naar centrale examens. De hogere status van centrale examens en handelingsverlegenheid van leraren daar waar het de constructie en beoordeling van schoolexamens betreft zijn daar in belangrijke mate debet aan. Ook het schoolexamen dient kwalitatief aan de maat te zijn, niet alleen in termen van validiteit maar ook waar het de betrouwbaarheid en transparantie betreft.

Meer investeren in de kwaliteit(sborging) van schoolexamens lijkt dus geboden. Om de kwaliteit van het schoolexamen op een hoger niveau te krijgen is het van belang leraren 'inspirerende steun in de vorm van specificaties en voorbeelden' (Kuiper, et al., 2013) te bieden bij het vormgeven van het schoolexamen. Dat kan door de in het schoolexamen te toetsen doelen al of niet webondersteund te voorzien van goede, niet-verplichtende uitwerkingen en aansprekende voorbeeldopgaven. Daarnaast is het aan te bevelen in regelgeving meer aandacht te hebben voor het specifieke karakter van het schoolexamen, bijvoorbeeld door te bepalen én er op toe te zien dat in het schoolexamen vooral of alleen

de niet aan het centraal examen toegewezen eindtermen aan de orde komen. Het centraal examen is immers niet normstellend voor de vorm en inhoud van het schoolexamen. Bij het verschaffen van ruggensteun kunnen ook regionale vaksteunpunten, vo-ho-netwerken en vakverenigingen een rol spelen, bijvoorbeeld door leraren van verschillende scholen te stimuleren schoolexamens te ontwikkelen en elkaars schoolexamens te bespreken onder begeleiding van toets- en vakexperts. Iets dergelijks zou ook in docentontwikkelteams (DOT's) georganiseerd kunnen worden. Ook kan gedacht worden aan onderlinge visitatie van scholen en aan een systematiek waarin een leraar van een andere school een rol krijgt bij de beoordeling van mondelinge of meer praktijkgerichte toetsing op een andere school. De rol van toetsexperts en inspectie is hier ook belangrijk. De inspectie zou toezicht kunnen houden op het kwaliteitssysteem (bv. RCEC en COTAN; zie ook BKE/SKE in mbo en hbo) en de transparantie daarvan.

Naast dit alles zullen ook (toekomstige) leraren zelf (meer) toegerust moeten worden op het punt van toets- en curriculumvaardigheden. Bij de scholing en nascholing van leraren zal er meer aandacht moeten komen voor het vertalen van op landelijk niveau geformuleerde doelen naar een meer valide keten van leerdoelen en inhouden, effectieve opdrachten/taken en toetsing, voor de onderlinge samenhang tussen leerdoelen, en dergelijke. Voor dit soort zaken is er op dit moment in de lerarenopleidingen vrij weinig aandacht. Ook zullen leraren ondersteund moeten worden bij de inzet van een ander toetsinstrumentarium. Voorbeelden van goede leerdoelen, passende taken en activiteiten, feedback- en correctiemodellen, rubrics en portfolio's kunnen leraren helpen om de gewenste kwaliteit te bereiken. Daarbij is de inzet onontbeerlijk van experts op het terrein van curriculumontwikkeling, toetsconstructie en (vak)didactiek. Er zouden toetsbanken kunnen worden ingericht, al of niet met ondersteuning van Cito of andere toetsspecialisten, zoals in het vmbo de schoolexamenbank met gevalideerde kennis- en praktijkopdrachten (items) voor de beroepsgerichte schoolexamens vmbo groen (zie www.degroenestandaard.nl/wat-is-schoolexamenbank). Deze schoolexamenbank is inmiddels ook beschikbaar voor de andere profielen. De toetsopgaven ervan worden aangeleverd door leraren aan een redactiegroep en een kwaliteitsgroep bewaakt de validiteit, betrouwbaarheid en bruikbaarheid. De kwaliteitsborging van het schoolexamen wordt zo door leraren in nauwe samenwerking met experts op het gebied van curriculum, toetsing en vakgebied vormgegeven. Het is belangrijk dat leraren leren omgaan met dat instrumentarium. Een goed beoordelingsmodel is belangrijk, maar ook belangrijk is dat leraren zoveel mogelijk op dezelfde manier met dat model omgaan (hetgeen tegelijkertijd enigszins op gespannen voet staat met gewenste flexibiliteit en eigenaarschap bij docent en school). Daar zou bij scholing en nascholing aandacht voor moeten zijn. Ook kan het collegiale overleg binnen de vakvereniging of binnen regionale steunpunten daaraan bijdragen.

4.3 Curriculaire uitdagingen

Het doelgericht vormgeven van onderwijs (curriculum) impliceert dat tijdens en aan het eind van de rit op een passende manier wordt nagegaan of die doelen worden gerealiseerd (formatieve en summatieve toetsing). Curriculum en toetsing vormen dus een twee-eenheid. Maar op die principiële verwevenheid valt in de praktijk wel wat af te dingen. Dat geldt zeker voor een brede en eigentijdse weerspiegeling van doelen en veranderingen daarin in toetsing en examinering. Een betere uitlijning in die zin is vanuit curriculumoptiek cruciaal en kan tevens als hefboom dienen voor de realisering van doelen en inhouden en van relevant geachte veranderingen daarin. Aldus kan de sterk sturende werking van regulering 'aan de achterkant' benut worden als hefboom voor inhoudelijke vernieuwing 'aan de voorkant' en in het onderwijsgebouw zelf, met dien verstande: toetsing volgt het curriculum, en niet omgekeerd. Noodzakelijk voor meer synergie tussen curriculum en toetsing zijn open communicatie en samenwerking tussen alle betrokken partijen, vanzelfsprekend met inachtneming van eenieders specifieke (stelsel)verantwoordelijkheid. Beide aspecten zijn randvoorwaardelijk voor het creëren van gemeenschappelijk draagvlak omtrent het nut en de noodzaak van een inhoudelijke vernieuwing (aanleidingen, motieven, uitgangspunten) en voor een cocreatieve uitwerking, beproeving en invulling daarvan. Nogmaals zij gewezen op het belang van een passende interpretatie en operationalisering van de vernieuwing in toetsing en examinering, met meer balans tussen validiteit, betrouwbaarheid en transparantie. Ook noodzakelijk is 'a well-defined sense of purpose' (Millar & Osborne, 1998), ofwel: houvast voor alle betrokkenen in de vorm van voldoende gespecificeerde doelen en inhouden, in combinatie met voorbeelduitwerkingen (zie ook Wallace & Priestley, 2017, p. 345).

In het voorgaande is toegelicht dat een curriculaire herbezinning op toetsing en ook verantwoording op zijn plaats is, in de context van Curriculum.nu maar ook los daarvan. Het adagium 'niet méér, maar anders en beter' is uiteengelegd in de thema's 'Kwaliteitsborging in termen van validiteit, betrouwbaarheid en transparantie', 'Meer formatief en minder summatief evalueren' en 'Meer schoolgebonden toetsing en centrale toetsing wat minder prominent'. Anders gezegd, ook los van Curriculum.nu is er alle reden na te denken over deze drie thema's, al biedt een dergelijk voornemen tot inhoudelijke herijking een uitgelezen mogelijkheid. Neem bijvoorbeeld de verhouding tussen ce en se. Uit dit duale systeem valt simpelweg meer te halen dan tot op heden gebeurt we nu doen. Daarbij kan het geen kwaad ons te realiseren dat er landen zijn (waaronder Canada, Australië, Vlaanderen en Ierland) die géén centrale examinering kennen, volstaan met heldere eindtermen (die richting en houvast verschaffen), veel voorbeelduitwerkingen aanreiken (die veel ruimte bieden voor eigen in- en aanvulling) en de bewaking van de

kwiteit van toetsing en examinering in handen leggen van de onderwijsinspectie. Deze constatering is louter bedoeld ter relativering van het belang van centrale examinering en zeker niet als pleidooi voor afschaffing ervan. Een onderwijssysteem zonder centrale afsluiting heeft immers ook duidelijke nadelen, met als belangrijkste dat het vervolgonderwijs alsnog in de rol wordt gemanoeuvreed van selectiefabriek.

De belangrijkste curriculaire uitdaging op het terrein van toetsing en verantwoording is het realiseren van meer balans tussen:

- de inhoudelijke strekking van examenprogramma's (die SLO ontwikkelt en beproeft) en de inhoud en vorm van centrale examens (ontwikkeld door Cito, onder regie van CvTE). Deze spanning komt naar voren in nagenoeg alle vakspecifieke trendanalyses (zie de samenvattingen daarvan elders in deze Curriculumspiegel). Vergelijkbaar is de spanning in het po tussen inhouden van vakken en uiteindelijke landelijke toetsen als lvs en (centrale) eindtoets. Vanuit SLO-perspectief is dit een punt van aanhoudende zorg, met de te geringe aandacht voor validiteit als rode draad. Dit alles raakt ook de beperkte aandacht in toetsing en verantwoording voor 'merkbare' doelen. Voortbouwend op eerdere projectactiviteiten (waaronder bijvoorbeeld 'Kwaliteitsborging se') en in samenspraak met andere direct betrokkenen wil SLO deze handschoen blijven oppakken.
- schoolgebonden en landelijke toetsing en examinering. Ook daarvan zijn hierboven en in de vakspecifieke bijdragen elders verschillende voorbeelden genoemd, van po tot en met tweede fase. De onbalans waarvan hier sprake is zit deels ook in verschillen in specificatie van eindtermen. Met ingang van 2007 zijn eindtermen in examenprogramma's geglobaliseerd om scholen meer ruimte te geven voor eigen keuzes. Echter, de eindtermen die zijn toegewezen aan het ce worden net als vóór 2007 onder regie van CvTE nader gespecificeerd in syllabi, daar waar de eindtermen die zijn gereserveerd voor het se meer open zijn geformuleerd en veel ruimte bieden voor eigen interpretaties. Voor het ce is daarmee per saldo de wijze van inhoudelijke regulering 'aan de voorkant' onveranderd gebleven. Overwogen zou kunnen worden alle eindtermen in examenprogramma's (enigszins) nader te specificeren alvorens die toe te wijzen aan ce en se. Dat, in combinatie met ruggensteun in de vorm van voorbeeldopgaven en dergelijke, zou scholen meer richting en houvast kunnen bieden bij het vormgeven van schoolexaminering.
- formatief evalueren en summatief evalueren. In het voorgaande is beargumenteerd dat er in beleid, praktijk en wetenschap duidelijk behoefte is aan aanvulling op de overwegend summatieve toetscultuur in het onderwijs in de vorm van meer aandacht voor formatief evalueren. Meer investeren in een feedbackcultuur sluit aan bij de behoefte aan meer differentiatie in leerroutes en methoden, meer flexibiliteit in inhoudelijke keuzes en beter maatwerk. Bovendien kan dat er voor zorgen dat minder

zwaar geleund hoeft te worden op summatieve toetsing (al blijft dat laatste nodig). Formatief evalueren draagt bij aan effectief onderwijs, maar doet tevens een stevig beroep op daartoe toegeruste en adequaat ondersteunde leraren en schoolleiders.

Daar waar het de professionalisering op het terrein van curriculum en toetsing betreft zijn nog werelden te winnen. In feite is dat een noodzakelijke voorwaarde voor alle inhoudelijk-didactische kwaliteitsverbeteringen van het onderwijs. Die professionalisering geschiedt bij voorkeur in velerlei vormen en op velerlei niveaus: in opleiding en nascholing; individueel en in teamverband; leraren en schoolleiders; scholen en educatieve partners daaromheen. Een eerste professionaliseringsthema betreft formatief evalueren (zie <http://curriculumvandetoekomst.slo.nl/projecten/formatieve-evaluatie>), met als focus de versterking van wat aangeduid zou kunnen worden als *pedagogical content, curriculum & assessment knowledge* (PCCAK). Ofwel, competentie-ontwikkeling op de combinatie van vakinhoud, vakdidactiek, curriculum en assessment. Bij ondersteuning kan gedacht worden aan (meer) houvast in de vorm van beschikbaarheid op landelijk niveau van heldere toetsbare leerdoelen (met uitdrukkelijke inachtneming van risico's als dichttimmeren, versnipperen en nog meer summatief toetsen), uit exemplarische uitwerkingen daarvan en uit uitwisseling en reflectie door middel van leernetwerken c.q. professionele leergemeenschappen ('voeden en gevoed worden'; zie de eerder genoemde leernetwerken voor vo en po). Ook internationaal zijn daar mooie voorbeelden van beschikbaar, onder andere uit Australië: voor verschillende vakken/domeinen - bijvoorbeeld *arts -dance* maar ook voor taal - zijn per leerjaar of koppels van leerjaren *achievement standards* opgesteld, voorzien van voorbeeldportfolio's met type opdrachten/taken/toetsen, uitwerkingen op, onder of boven niveau, en annotaties (<http://www.australiancurriculum.edu.au>). In andere landen hebben scholen de beschikking over websites en *guidances* met veel praktische handvatten. Voorbeelden daarvan zijn Ierland (<http://ccea.org.uk>) en Canada (<http://www.edugains.ca>).

Een tweede belangrijk professionaliseringsthema is de kwaliteitsborging van het se. In deze thematiek is door SLO de afgelopen jaren behoorlijk geïnvesteerd (handreikingschoolexamens.slo.nl). De producten en diensten die dat alles heeft opgeleverd voor het vmbo en de tweede fase vragen om doorontwikkeling, verdere verspreiding en benutting, ook voor professionaliseringsdoeleinden. Zie in dit verband ook de beschikbaarheid van de schoolexamenbank voor de verschillende vmbo-profielen.

Referenties

Adie, L., & Willis, J. (2016). Making meaning of assessment policy in Australia through teacher assessment conversations. In D. Laveault & L. Allal (eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 35-53). Springer International Publishing Switzerland.

Assessment Reform Group (1999). *Beyond the black box*. Cambridge: University of Cambridge School of Education.

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. London: NFER/Nelson.

Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2003). *Assessment for learning. Putting it into practice*. Maidenhead: Open University Press.

Brookhart, S.M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3-12.

CCEA (2009). *Assessment for learning. A practical guide*. Belfast: Council for the Curriculum Examinations and Assessment.

DeLuca, C. LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251-272.

Emons, W.H.M., Glas, C.A.W., & Berding-Oldersma, P.K. (2016). *Rapportage vergelijkbaarheid eindtoetsen*. Utrecht: Expertgroep po.

Gulikers, J. & Baartman, L. (2017). Doelgericht professionaliseren: Formatieve toetspraktijken met effect! Wat DOET de docent in de klas? (conceptversie). Den Haag: NRO-PPO.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational research*, 77(1), 81-112.

Inspectie van het Onderwijs (2010). *De staat van het onderwijs. Onderwijsverslag 2008/2009*. Utrecht: Inspectie van het Onderwijs.