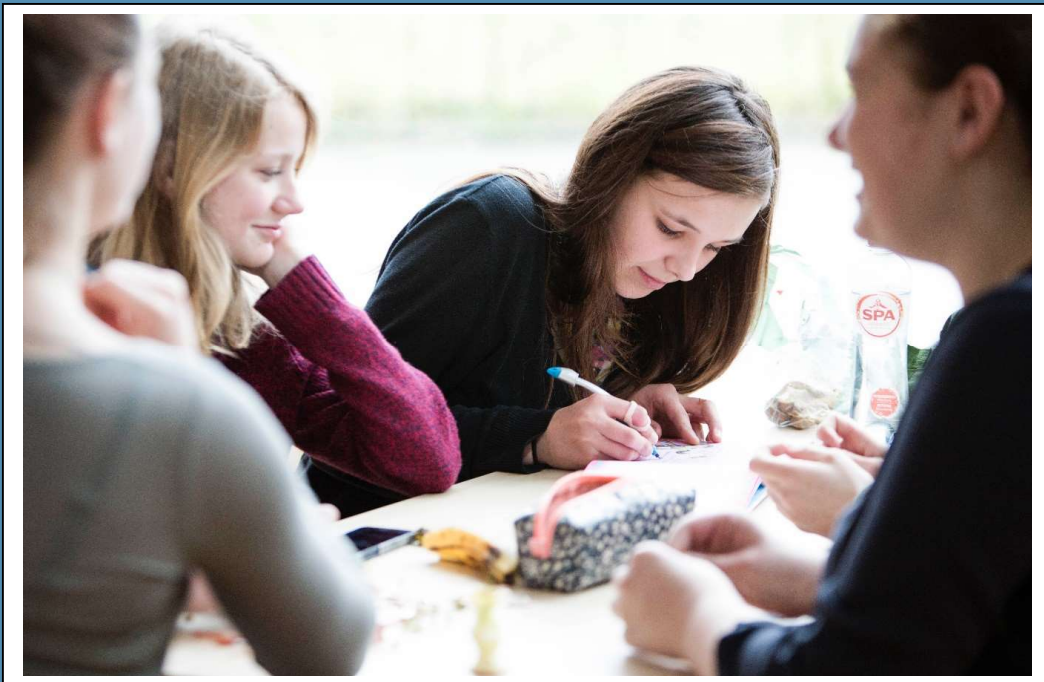
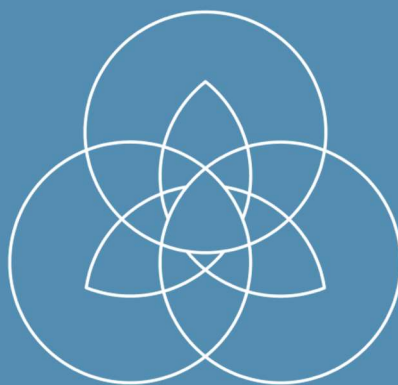




Achter elk cijfer schuilt een uniek verhaal

Praktijkvoorbeeld van een kiemkrachtig beoordelingsproces



Ariëlla Krijger & Wouter Modderkolk

Lectoraat Waarde(n) van Vrijeschoolonderwijs

Achter elk cijfer schuilt een uniek verhaal

Inleiding

Sinds de invoering van het centraal examen in de vrijescholen voor voortgezet onderwijs (Project 2000) staan we voor de uitdaging om juist ook bij de uitvoering van het voorgeschreven examencurriculum een brede ontwikkeling van de individuele leerling aan te moedigen. De ervaring in de school laat zien dat leerlingen goed willen scoren op de onderdelen van het schoolexamen zoals beschreven in het programma van toetsing en afsluiting (afgekort PTA) (Modderkolk, 2017). Dit doet een extra appel op het pedagogisch kunstenaarschap van leraren want volgens de eigen intenties van vrijeschoolonderwijs richt een beoordeling zich op méér dan cijferwaardering alleen. Leraren richten zich niet alleen op het geheugen en verstand maar nemen voor alles waar wat pedagogisch gezien in de opgroeiende jonge mens aan het ontkiemen is in directe relatie tot het leven zelf op grond van echte ervaringen (Steiner, 1996). Daarmee staan de leraren voor de uitdaging om de lessen en toetsen zodanig in te richten dat ze niet alleen bijdragen aan examensucces maar vooral aan menswording in brede zin. Een van de kernwaarden van vrijeschoolonderwijs daartoe is dat het curriculum *middel* tot ontwikkeling is en geen doel op zich vormt.

Dit onderzoek richt zich op de afname en beoordeling van een toets die meetelt voor het schoolexamen en die specifiek ontworpen is om ook ruimte te geven aan deze menswording in brede zin. We zoeken daarom naar een natuurlijk evenwicht tussen een ontwikkelingsgerichte, *kwalitatieve* beoordeling door het geven van individuele feedback en een examengerichte, *kwantitatieve* beoordeling door de becijfering van leerlingenprestaties op een schaal van 10 volgens door de overheid genormeerde standaarden. We ontwerpen en testen daartoe een beoordelingsmodel dat deze beide benaderingen – ontwikkelingsgericht en examengericht – dienstbaar maakt aan de beoogd brede ontwikkeling van onze leerlingen. We staan in dit streven om meer werkvormen met een pedagogisch oogmerk in de onderwijspraktijk op te nemen bepaald niet alleen. Talrijk onderzoek van de laatste decennia laat zien dat het geven van feedback en toepassingen van diverse dialogische onderwijsvormen een positieve werking hebben op het zelfvertrouwen van leerlingen hetgeen aantoonbaar leidt tot verbeterd leren (Black & Wiliam, 1998; Modderkolk, 2017). Deze *formatieve beoordelingspraktijk* werkt *vormend* op de brede ontwikkeling van leerlingen. Daarnaast staat de *summatieve beoordeling* die op basis van de resultaten tot een becijferde slotsom komt. Uit veel onderzoek is naar voren gekomen dat bij de *combinatie* van summatieve en formatieve beoordelingen de leerlingen bij de teruggave van een toets vooral letten op de rechtvaardigheid van het (hopelijk nog onderhandelbare) cijfer terwijl de eveneens gegeven schriftelijke feedback nauwelijks wordt opgemerkt (Black, Harrison, Lee, & Wiliam, 2004; Wiliam, 2012). Een belangrijk leermoment verdwijnt hiermee naar de achtergrond.

Omdat het geven van cijfers voor het schoolexamen verplicht is en tegelijk de zojuist aangehaalde brede ontwikkeling van leerlingen door alle vakken heen centraal staat in de vrijeschool, hebben de auteurs van dit artikel zichzelf de volgende vraag gesteld:

Hoe kun je het verhaal dat achter een cijfer schuilgaat zichtbaar en inzichtelijk maken opdat het vuur van enthousiasme dat leerlingen met al hun vermogens aanzet om van het leren en (later) van het leven zelf te willen leren, niet uitdooft maar brandende blijft? Wat is binnen deze context een kiemkrachtig oordeel?

Om tot beantwoording van deze vraag te komen volgen we nauwgezet de praktijk van een toets waarbij leerlingen van de elfde klas een essay schrijven voor het vak economie. We geven een overzicht van de verschillende fasen tijdens het onderzoek betreffende 1) opdracht, 2) beoordelingscriteria en beoordelingsproces, 3) deliberatie over de bevindingen, 4) reactie van de leerlingen bij de teruggave van het werk, en 5) leerlingenevaluatie over de beoordeling van hun essay. Met wat uit de verzamelde gegevens naar voren komt, willen we de volgende twee sub-vragen beantwoorden:

- a. *Hoe kun je de samenhang tussen het unieke verhaal van de individuele leerling en het toegekende cijfer betrouwbaar en navolgbaar maken doordat dit summatief toetsbaar is voor externe betrouwbaarheid en intern formatief werkt binnen de onderwijspedagogische relatie van leraar en leerling?*
- b. *Op welke wijze draagt deze praktijk er toe bij dat leerlingen medeauteurschap van de eigen beoordeling op zich nemen?*

Dit actieonderzoek maakt gebruik van fenomenologische (en kritische) onderzoeksstrategieën (Ponte, 2012; Rawson, 2010) om de antwoorden op de twee sub-vragen te vinden waarmee we vervolgens de kernvraag beantwoorden.

De aanleiding voor de opdracht en de opdracht

“Economie gaat niet over rekenen en wiskunde (modellen zijn slechts uitdrukking van een verhaal en een poging om de wereld om ons heen rationeel te begrijpen), maar over verhalen, mythen en archetypes, over mensen met emoties én verstand en de goede en slechte zaken die mensen doen en de keuzes die zij maken.

Het denken over economie is al erg oud, zelfs in een van de oudste literaire werken (het Gilgamesj-epos) kunnen we gedachten vinden die raken aan ons huidige denken (Sedláček, 2012). Dit helpt ons het huidige economische denken te begrijpen. Grieken zoals Plato, Aristoteles en Socrates hebben belangrijke fundamenteën gelegd waar ons denken uit voort komt. Ons economische denken is sterk beïnvloed door de filosofie, godsdienst en had vroeger altijd een ethische component in zich. In de 20ste eeuw is het economische denken sterk beïnvloed door het determinisme en het wiskundig rationalisme waardoor we vandaag de dag vooral formules en grafieken associëren met economie.

Toch blijft economie een sociale wetenschap (het gaat immers om het menselijk streven naar welvaart) en het kan gevaarlijk zijn als wij de economie te vaak als exacte wetenschap behandelen. Daardoor kunnen wij het verhaal achter cijfers, modellen, formules en grafieken (en de ethiek) uit het oog verliezen.”

Uit: de opdracht zoals gegeven aan de leerlingen door Modderkolk (2015)

Aan de leerlingen wordt gevraagd om een essay te schrijven waarin zij een standpunt innemen over een zelfgekozen casus. De opdracht om een essay te schrijven telt mee voor het schoolexamen (SE). Het resultaat vormt 8% van het SE-cijfer voor economie. Het essay bevat tussen de 1500 en 2000 woorden (exclusief citaten en bijlagen) en de opbouw van het essay (zinsbouw, spelling) volgt hetgeen aangeleerd is bij het vak Nederlands. De leerling dient in een bijlage (beknopt) aan te geven welk aspect uit de literatuur is verwerkt en de bronnen dienen volgens de APA-richtlijnen te worden vermeld.

De leerlingen moeten de casus vanuit verschillende invalshoeken te bestuderen. Zij maken in ieder geval gebruik van de volgende bronnen:

- Bregman & Frederik (2016). *Waarom vuilnismannen meer verdienen dan bankiers*.
- Tomáš Sedláček (2012). *De economie van goed en kwaad*: hoofdstuk 13..
- Polak (2008). *Game theory; Open lessons Yale University*. Lessenreeks over speltheorie, bewerkt en vertaald naar het Nederlands door Wouter Modderkolk.

Uit deze literatuur nemen de leerlingen per bron een of meer aspecten die hen aanzetten tot nadenken. De gekozen aspecten uit de literatuur moeten door een eigen denkproces met elkaar in verband gebracht worden en de leerlingen dienen hierover een standpunt in te nemen dat gebaseerd is op een bredere kijk van het economische denken. Met andere woorden: zij dienen aan te geven waarom een bepaalde ontwikkeling goed is of juist slecht en hoe de economische orde er in hun ogen uit zou moeten zien.

Beoordelingsproces en beoordelingscriteria

Het essay zal gebruikmakend van 'rubrics' (Jonsson & Svingby, 2007) gewaardeerd worden op de volgende beoordelingscriteria:

- a) Vorm (weging 1x): opbouw en bronvermelding.
- b) Inhoud (weging 2x): literatuurkoppeling met de casus en de juistheid van de argumenten
- c) Eigenheid (weging 2x): standpuntbepaling en originaliteit.

Belangrijk doel van deze opdracht is dat we bij de beoordeling rekening houden met alles wat pedagogisch gezien in de opgroeiende jonge mens aan het ontkiemen is (Steiner, 1996). We gaan uit van een brede opvatting van onderwijs die in onze tijd door onderwijspedagoog Gert Biesta bijvoorbeeld beschreven is in drie met elkaar samenhangende doeldomeinen: *kwalificatie*, *socialisatie* en *subjectwording* (2012, 2015). Lector Aziza Mayo (2015, 2019) heeft deze drie begrippen verbonden met waarden van het vrijeschoolonderwijs. Zij kenmerkt het vrijeschoolonderwijs als 'subjectiverend onderwijs', waarbij de kwalificatie en socialisatiefuncties steeds invulling krijgen in het licht van de subjectwordingsfunctie (zie ook Biesta, 2018).

Vanuit dit perspectief op onderwijs als breed en vormend, staat *kwalificatie* vooral in relatie tot wat vakinhoudelijk gevraagd wordt. Leerlingen moeten zich niet alleen de stof hebben eigengemaakt, maar boven de stof staan om dit essay te kunnen schrijven. *Socialisatie* staat hier in relatie tot het vermogen van leerlingen om mede-eigenaar te zijn van een formatief onderwijsproces; het vraagt bijvoorbeeld ook om het zich kunnen invoegen in cultureel bepaalde vormvereisten m.b.t. opbouw en bronvermelding. De leerling verschijnt (in eigenheid) als *wordend subject* als het betoog vindingrijk is qua standpuntbepaling en origineel is in de beantwoording van vragen en het opstellen van (vervolg)vragen die uit de casus opkomen, waarbij ethische aspecten niet uit de weg worden gegaan.

Deze brede benadering van onderwijs vergt bij de beoordeling van wat leerlingen bereikt hebben een specifiek op het individu gerichte benadering in relatie tot - maar idealiter ook los gezien van - de vereiste standaarden. Om daaraan tegemoet te komen krijgen de leerlingen naast het op rubrics-scores gebaseerde cijfer aanvullend een geschreven beoordeling die het verhaal beschrijft dat achter de becijfering schuilgaat. Ook de leerlingen zelf dragen hieraan bij. Zij schrijven ongeveer vier weken na het inleveren van hun essay een zelfbeoordeling. De economieleraar laat ze daartoe bewust pas na een langere periode - dus met voldoende afstand - nog eens kritisch naar hun werk kijken. In deze zelfbeoordeling wordt aan de leerlingen per beoordelingscriterium gevraagd in

welke mate zij er in zijn geslaagd om aan de criteria te voldoen. Vervolgens wordt hen gevraagd om één aspect te formuleren dat ze een volgende keer anders willen doen. De leraar benadrukt daarbij dat een reële zelfbeoordeling wordt meegewogen bij de definitieve beoordeling. Het beoordelingsformat voorziet in ruimte voor een iets hoger cijfer vanwege bijzondere prestaties.

De betrouwbaarheid en navolgbaarheid van een kwalitatieve beoordeling van de essays wordt nagestreefd door zowel het gebruik van een *beoordelingsrubric* (Jonsson & Svingby, 2007) als door de toepassing van het *principe van triangulatie* (Mortelmans, 2007). Dat wil zeggen dat de essays vanuit meerdere perspectieven worden beoordeeld, in dit geval dat van de docent, de leerling en de onderzoeker. De beoordelingsrubric met de criteria vorm, inhoud en eigenheid waardeert de diverse aspecten met: onvoldoende (0 punten), voldoende (1 punt) of goed (2 punten). Wegingsfactoren en waarderingsfactoren leiden samen tot het cijfer waar alleen een bijzondere prestatie als bijvoorbeeld de zelfbeoordeling van de leerling nog invloed op kan uitoefenen.

De leerlingen hebben vooraf uitgebreide en ruim toegelichte informatie ontvangen over de opdracht, de beoordelingscriteria en het beoordelingsproces. Deze informatie is besproken in de les.

Deliberatie over de bevindingen

In dit deel komt samen wat het betekent om aan de hand van beoordelingsrubrics en (zelf)beoordelingen vanuit drie perspectieven (leraar, onderzoeker, leerling) een totaal oordeel vast te stellen bestaande uit SE-cijfer en feedbacktekst. Op deze plaats is het van belang om de verschillende achtergronden toe te lichten van waaruit de auteurs, Wouter Modderkolk (docent/onderzoeker) en Ariëlla Krijger (onderzoeker) hun waarnemingen doen. Modderkolk is de economiedocent van deze 11^e klas. Van hem komt de gehele opdracht inclusief de voor het SE vereiste beoordelingscriteria. Tijdens een masterstudie heeft hij zich specifiek gericht op uiteenlopende vormen van dialogisch onderwijs m.n. door het geven en ontvangen van feed up, feedback en feed forward (Modderkolk, 2017). Oud-vrijeschoolleraar Krijger heeft geen economische achtergrond. Wel heeft zij sinds het begin van haar researchmaster (2009-2013) in diverse landen en omstandigheden onderzoek gedaan naar wat vrijeschoolleerlingen (kunnen) laten zien op het gebied van criteriagerichte reflectie en zelfevaluatie bij de waarneming van het eigen werk. Zij stuitte daarbij op het *contemplatieve* karakter van deze speciale vorm van reflectie (Krijger, 2016). De samenwerking tussen beide onderzoekers is dus geen optelsom van twee ongeveer dezelfde achtergronden, juist de aanvullende waarde van verschillende bekwaamheden vormde een uitdaging om samen dit project te doen.

We hebben los van elkaar de essays en de leerlingen-zelfbeoordelingen gelezen. Na een intens waarnemingsproces werd het rubrics-format door beiden apart ingevuld, voorzien van een cijfer en van feedback over *hoe* de leerling in het essay in brede zin naar voren is gekomen. Dit proces is volgens een narratieve onderzoeksmethode uitgevoerd. Het waarnemingsproces is uitgegaan van een combinatie van diepgaand en meditatief lezen van de essays dat leidde tot een interpretatie en selectie van betekenisvolle fragmenten (Clandinin & Connelly, 2000). Vervolgens bespraken we elk essay afzonderlijk tijdens een drie uur durend overleg. We vergeleken cijfers, we lazen elkaar (fragmenten van) feedbackteksten voor en zochten - zeker bij uiteenlopende waarnemingen - naar consensus voor een afgewogen, fair oordeel. Als we inzoomen op de verschillende onderdelen van de rubrics, de leerling-zelfbeoordeling en onze feedback ontstaat het volgende beeld:

Rubrics, de route naar een cijfer

De verschillen tussen de cijferwaarderingen die we los van elkaar gaven vertellen een eigen verhaal over onze inschattingen van de prestaties van leerlingen bij het gebruik van rubrics. Opmerkelijk is dat de meeste *cijferwaarderingen* dicht bij elkaar lagen, maar liefst 74% van de beoordelingen verschilden één punt of minder en bij 48% was het verschil slechts een halve punt (N=23). We hebben een sterk vermoeden dat het gebruik van rubrics hieraan heeft bijgedragen.

Bij de analyse van de verschillen stuiten we op twee opvallende fenomenen:

Allereerst was de *formulering van de indicatoren* voor evaluatiecriteria niet altijd eenduidig. Waar moet een argument aan voldoen om helder en juist te zijn? Wanneer is een bron op juiste wijze gekoppeld aan de casus? Door dit bij een volgende gelegenheid aan te scherpen zouden de verschillen in waardering nog verder terug te brengen zijn (Arter & Chappuis, 2006). Ook Jonsson en Svingby benadrukken het belang van een effectief ontwerp en een goed begrip en gebruik van de rubrics. Dit zou cruciaal zijn voor de betrouwbaarheid en navolgbaarheid van deze beoordelingen (2007).

Ten tweede had Krijger als tweede beoordelaar – niet ‘gehinderd’ door *vakspecifieke kennis* – het voordeel van de onbevangen blik op de *inventiviteit* van de wijze waarop een leerling argumenten ter tafel brengt. Zij keek naar het geheel, naar de naleving van vormvoorschriften, de eigenheid van het standpunt, en *hoe* uit het geheel originaliteit naar voren kwam. Zij nam daarom vooral waar wat zich afspeelde binnen de doeldomeinen socialisatie en subjectwording. Markant was dat onaanvaardbare verschillen van 1,5 punt of meer voortkwamen uit echt vakspecifieke fouten die Modderkolk als vakleraar en eerste beoordelaar direct opmerkte. We ontdekten dat een te grote nadruk op vakinhoudelijke onjuistheden en de naleving van vormvoorschriften, m.a.w. op de kwalificatie- en socialisatiekant van het onderwijsproces, wel eens een ‘blinde vlek’ veroorzaken kan voor subjectwordingsaspecten als *inventiviteit* en *uniciteit* die bij een brede benadering zeker ook moeten meewegen in het oordeel.

De leerlingen-zelfbeoordeling

Het andere aspect dat bij de deliberatie naar voren kwam was de leerlingen-zelfbeoordeling. Deze werd geschreven aan de hand van vragen over naleving van vormvereisten uit de opdracht, de koppeling van de drie bronnen aan de casus, de helderheid van de argumenten in het betoog leidend tot een door de literatuur onderbouwd standpunt waaruit originaliteit naar voren komt. Tenslotte: wat zou je een volgende keer anders doen? Bij de zelfbeoordeling van leerlingen keken we vooral naar het zelfkritisch vermogen op basis van waarnemen en denken. In de antwoorden kwam vaak naar voren dat leerlingen in het vervolg eerder met zo’n opdracht zouden beginnen, dat ze hun werk nog eens kritisch door een ander zouden laten lezen en dat ze meer aandacht aan zinsbouw enz. wilden besteden. Dit zijn zaken die vooral de planning en organisatie van het werk betreffen. Maar ook kon je lezen:

"Bij de eerste vraag heb ik gezegd dat ik aan het einde te veel ga herhalen. Dat zou ik de volgende keer dan ook anders doen. Ik moet de volgende keer zorgen voor meer onderbouwingen. Dus langer denken voordat ik echt iets opschrijf, meer research doen zodat ik betere/ meer onderbouwingen heb en dus niet in herhaling val".

"Ik denk dat ik een volgende keer mijn betoog duidelijker naar voren laat komen; en deze steviger zal onderbouwen met standpunten. Daarnaast zal ik aan mijn essay mijn eigen mening toevoegen om het wat interessanter en aantrekkelijker voor een potentiële lezer te maken."

Dergelijke inhoudelijke reflecties op het eigen werk kwamen in aanmerking voor een extra waardering. In totaal hebben we tien extra waarderingen gegeven vanwege bijzondere prestaties. Vijf daarvan betroffen de *kwaliteit van de zelfbeoordeling*. Het kwam zelfs voor dat een essay op zich genomen niet zo sterk uit de verf kwam maar dat de zelfbeoordeling zo trefzeker was dat de beoordelaars nauwelijks iets hadden toe te voegen. Deze leerling heeft dan na die weken van loslaten en opnieuw focussen toch nog veel bereikt.

Bij de overige vijf extra waarderingen ging het in drie gevallen over de kwaliteit en diepgang van het essay, één betrof de originaliteit en één de schrijfstijl en structuur van het stuk.

De cijferwaarderingen werden vastgesteld aan de hand van de rubrics-resultaten en de leerlingen-zelfbeoordelingen. Van de in het totaal 23 cijferwaarderingen is in tien gevallen de waardering gelijk gebleven, in acht gevallen werd het cijfer met een half punt verhoogd, driemaal met een half punt verlaagd en tweemaal is het cijfer met één punt omhooggegaan. Tijdens de deliberatie werd elk besluit hiertoe steeds genomen door Modderkolk als verantwoordelijk docent.

De feedbacktekst van beide onderzoekers

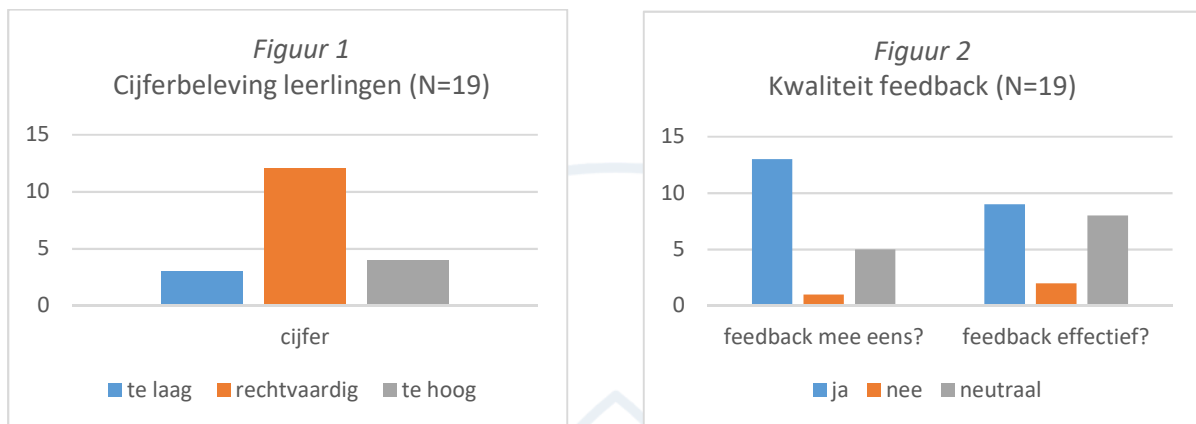
Tijdens de deliberatie over de geschreven beoordelingen en de accenten die elk van de onderzoekers legden in hun waardering van de essays, werden we ons in toenemende mate bewust van de aanvullende waarde van beide benaderingen. Aspecten uit de teksten van Krijger met aanvullende, verbredende feedback zijn daarom opgenomen in de teksten van Modderkolk.

Reactie van de leerlingen bij de teruggave van het werk

Toen de leerlingen hun werk terugkregen viel het de docent op dat de leerlingen in stilte hun beoordelingsdocument doorlazen. Ze waren nieuwsgierig naar de feedback die zij hadden gekregen. En vroegen ook aan elkaar welke opmerkingen, complimenten en adviezen zij hadden gekregen. Een aantal leerlingen kwam naar de docent toe om de feedback te bespreken. Het was opvallend dat er nauwelijks commentaar was op het gegeven cijfer. Leerlingen gaven aan dat zij de opdracht als zeer waardevol ervoeren. Een aantal leerlingen heeft het essay – na verwerking van de feedback – ook weer gebruikt bij het eindwerkstuk. Tijdens de lectoraatsdag van de Hogeschool Leiden (juni 2018) heeft een leerling tijdens een werkgroep van de auteurs over haar ervaringen met deze opdracht verteld. Het was van grote toegevoegde waarde dat ook een leerling – om wie het tenslotte gaat – haar perspectief op de opdracht deelde en op die manier deelnam aan de onderwijsdiscussie over breder toetsen en beoordelen.

De leerlingevaluaties over de beoordeling van hun essay

Vervolgens hebben we de leerlingen gevraagd naar hun mening over de opdracht. Een week later gaven ze feedback op het hele proces vanaf het ontvangen van de opdracht, de uitvoering ervan, het schrijven van een zelfbeoordeling tot en met het terugkrijgen van het schriftelijk beoordeelde en becijferde werk. Zij beantwoordden daartoe drie gesloten en drie open vragen. De eerste twee gesloten vragen betroffen de hoogte van het cijfer en de kwaliteit van de feedback. De derde vraag richtte zich op de *effectiviteit* van de feedback zodat de leerling helder inziet wat hij een volgende keer beter zou kunnen doen. Met de open vragen wilden we te weten komen 1) wat ze over zichzelf en hun werk hebben geleerd door de beoordeling, 2) wat zij in een ideale beoordeling zouden willen lezen en 3) wat ze van de opdracht vonden, of ze meer hebben geleerd dan economie alleen en of er verbeteringsuggesties zijn.



Figuur 1 en 2 vormen een weergave van wat uit de drie gesloten vragen naar voren kwam. Bij de vraag naar de effectiviteit van de feedback was de score tussen 'ja' en 'neutraal' bijna gelijk. Betekent dit dat leerlingen voor neutraal in de betekenis van 'veilig' kiezen omdat ze een aarzeling voelen over de inhoud van de feedback zelf of dat ze zich afvragen welke feedback effectief is en welke niet of minder? Dit zou nader onderzocht kunnen worden.

Opvallende feedback over wat leerlingen aan de beoordeling zeggen te hebben gehad (vraag 1), is *de behoefte om het beter beargumenteerd leren onderbouwen van een standpunt en een nauwlettender omgang met spelling en bronvermelding*. In de ideale beoordeling (vraag 2) willen veel leerlingen *concrete tips en tops*, sommigen vragen ook om een duidelijke balans tussen die twee waardoor ze hun cijfer beter kunnen begrijpen. Een enkeling vroeg naar *een persoonsgerichte beoordeling los van de vereiste niveaus*. Deze leerling wil – zo begrijpen we – weten of hij in de *eigen* ontwikkeling vooruitgegaan is vergeleken bij vorige prestaties.

In antwoord op de laatste open vraag werd de opdracht door een aantal leerlingen expliciet als goed en leerzaam ervaren met nadruk op *het leren uitvoeren van een onderzoek aan de hand van een eigen casus*. *Leren argumenteren en komen tot een eigen standpunt* werd ook waardevol gevonden evenals *de omgang met vormvoorschriften*. Niet alle leerlingen vonden het even gemakkelijk. De omgang met enerzijds een *vrije* opdracht en anderzijds met *onvrije* vormvoorschriften riep ook weerstand op. Over het algemeen genomen sprak voldoening uit de evaluaties: een goede opdracht, ervaring kunnen opdoen met een werkstuk, leerzaam, goed voorbereid en goed aangepakt.

Werking

De hele opdracht, inclusief de zelfreflectie en de uitvoerige feedback die de leerlingen kregen, is zowel door de leerlingen als door de docent als zeer waardevol ervaren. Leerlingen ervoeren dat hun werk serieus werd genomen en stelden de brede benadering bij de beoordeling van hun werk op prijs. Een aantal van hen heeft zelfs gevraagd of ze nog een dergelijke opdracht mogen doen, omdat ze er zoveel van hebben geleerd. En dan niet alleen op het gebied van het schrijven van een essay maar ook op metacognitief gebied.

Het was opvallend hoe waardevol het bleek om de leerlingen na een bepaalde rusttijd weer te bevragen over hun eigen werk. Sommige leerlingen gaven aan dat zij dit als confronterend

ervoeren, omdat zij geconfronteerd werden met hun eigen fouten of slordigheden. Juist die confrontatie met het eigen werk leidde tot een rijke reflectie. De vragen waren zo geconstrueerd (zie de paragraaf over de zelfbeoordeling van leerlingen op pagina 4 e.v.) dat zij vanuit een positieve insteek hun eigen werk gingen beoordelen. Leerlingen gingen door deze manier van vragenstellen niet *onnodig* kritisch naar het eigen werk kijken.

Leerlingen leken oprecht geïnteresseerd in het verhaal achter een cijfer. Diverse leerlingen kwamen naar de docent omdat zij bepaalde opmerkingen in de beoordeling niet geheel begrepen. Het verhaal achter het cijfer leek dus belangrijker dan het cijfer zelf. Er waren ook enkele leerlingen die vroegen naar een *ipsatief*¹ oordeel. Dat wil zeggen dat zij wilden weten in hoeverre ze *zelf* - ten opzichte van vergelijkbare prestaties in het verleden - een volgende stap hebben gezet. Het is voor een vakdocent lastig om dit te expliciteren omdat hij in de regel onvoldoende op de hoogte is van de kwaliteit van eerder geschreven essays bij andere vakken. Voor een aantal leerlingen is met de leraar Nederlands overleg gevoerd om aan de wens van deze ipsatieve-feedback gehoor te geven. Nader onderzoek kan uitwijzen hoe we leerlingen naast formatieve en summatieve ook meer ipsatieve feedback kunnen geven. Gedacht zou kunnen worden aan het werken met een portfolio (Koch, 2010).

Wat kunnen we leren van dit onderzoek?

In antwoord op subvraag a) komt in dit onderzoek naar voren dat het mogelijk is om een summatieve (resultaatgerichte) en formatieve (leerprocesgerichte) beoordeling te geven voor de uitvoering van een en dezelfde opdracht. Het summatieve gedeelte zit er in dat leerlingen een cijfer krijgen dat meetelt voor het schoolexamen. Het formatieve aspect van de beoordeling komt naar voren in hun eigen beoordeling en de feedback die zij krijgen van de beoordelaars. De impact van wat de leerlingen verworven hebben wordt versterkt als de beoordeling van deze opdracht weer wordt gebruikt als startpunt voor het schrijven van een nieuw essay, bijvoorbeeld bij het vak Nederlands (Wiliam, 2012).

Eveneens komt naar voren dat om tot een zo valide en betrouwbaar mogelijke, summatieve (cijfer)beoordeling te komen, de rubrics eenduidig(er) moeten worden geformuleerd. De verschillen in waardering – tussen Modderkolk en Krijger – maakten tijdens de deliberatie zichtbaar in hoeverre de beoordelingscriteria nog specifieker beschreven kunnen worden. Allen & Tanner (2006) wijzen er op dat waardeoordelen zoals goed, voldoende, onvoldoende, creatief, leuk e.d. moeten worden vermeden. Het is beter om te beschrijven wanneer een werkstuk als uitstekend (= een tien) wordt beoordeeld.

¹ Ipsatief komt van 'ipse' (= zelf Lat.). Een ipsatieve beoordeling spreekt zich uit over de voortgang van de individuele leerling *zelf*. Deze aandacht voor de individuele ontwikkeling is wat de kern uitmaakt van de getuigschriftbeschrijvingen in het vrijeschoolonderwijs (Krijger, 2018; zie ook de achtste voordracht van de *Oxfordcyclus* van Steiner (2004)).

Leerlingen gaven achteraf aan dat zij een dergelijke opdracht als leerzaam hebben ervaren vanwege de eigenheid die zij in deze opdracht kunnen leggen. De eigenheid komt hierin niet alleen terug in de opdracht zelf, maar ook in het feit dat de leerling meesprekt in de beoordeling. Hierdoor worden leerlingen ook uitgenodigd om te reflecteren op de weerstand die zij mogelijkwijs hebben ervaren tijdens het schrijven van hun eigen essay. Biesta (2018, p. 43) stelt dat onderwijs gaat over omgaan met weerstand. Niet zozeer dat we leerlingen moeten laten omgaan met weerstand door het ze maar moeilijk te maken, maar omdat volwassen in de wereld staan niet om weerstand heen kan gaan.

Voor de toekomst

Uit ons onderzoek komen een aantal punten naar voren die we zeker meenemen voor de toekomst. Allereerst willen we kijken in welke mate we de leerling ook mede-eigenaar kunnen maken van de formulering van beoordelingscriteria (subvraag b). Wat gedaan zou kunnen worden is dat leerlingen een aantal essays van het vorige jaar lezen en dat we op basis van die - door de docent als goed beoordeelde - essays samen komen tot beoordelingscriteria. Doordat deze beoordelingscriteria samen zijn opgesteld weten de leerlingen beter wat er van hen verwacht wordt (Rawson, 2015) en ontwikkelen ze een neus voor kwaliteit (Sadler, 1989). Het kan zelfs interessant zijn om de leerlingen samen een rubric te laten invullen met succescriteria. Dit draagt wel het risico in zich dat leerlingen dan (veiligheidshalve) gaan toewerken naar vooral objectiveerbare succescriteria (Straetmans, 2006) terwijl de kwaliteit van een essay zeker niet volledig in objectieve criteria te vangen is. Dominique Sluijsmans (2019) waarschuwt daarom ook voor het overmatig gebruik van rubrics. Zij geeft aan dat het soms verstandiger kan zijn om gebruik te maken van een single-point-rubric waarbij alleen een criterium voor een voldoende wordt geëxpliciteerd en met leerlingen naar goede voorbeelden wordt gekeken om excellentie te duiden.

Wat niet als vraag geformuleerd was, maar wel blijkt uit dit artikel is dat deze opdracht nogal arbeidsintensief was voor zowel de leerling (het creatie-, uitvoerings- en evaluatieproces) als de leraar en onderzoeker (de uitgebreide beoordeling). Een dergelijke, tijdrovende opzet door twee beoordelaars en de leerling zelf is niet zondermeer haalbaar. Het is de moeite waard om te onderzoeken met leraren van andere vakken of een dergelijke opdracht mede kan worden nagekeken door bijvoorbeeld een leraar Nederlands. Op deze manier blijft het concept van triangulatie (Mortelmans, 2007) voor een betrouwbaar kwalitatief oordeel behouden en ook het daarmee samengaande, als zeer vruchtbaar ervaren intervisieaspect. Daarnaast zou het waardevol zijn als een leerling dit essay kan plaatsen in een rij van vergelijkbare, grote opdrachten. Dit geeft hen ruimte om vakoverstijgende vaardigheden – lees hier: zelfstandige oordeelsvorming – te ontwikkelen met betrekking tot argumentatie, betoog en standpuntbepaling. Het geeft de leraar ruimte om ook ipsatieve feedback te integreren in zijn beoordelingspraktijk met het oog op de ontwikkelingstappen van de leerling *zelf* in vergelijking met een voorgaande resultaten. Dat leerlingen hier zelf ook scherp zicht op hebben is in dit praktijkvoorbeeld onmiskenbaar naar voren gekomen.

Ten slotte

Vanuit dit onderzoek zou het idee kunnen ontstaan dat we dergelijke opdrachten met regelmaat in ons curriculum moeten opnemen. Waar het gaat om reflectievaardigheden van leerlingen past hier

wel enige wakkerheid. Reflectie doet een appel op vermogens tot waarneming, vergelijking en beoordeling die om ontwikkeling en begeleiding vragen. Niet zelden is er sprake van een overdaad aan reflectieopdrachten die veelal tot een routinematige aanpak leidt. Dan zuigen leerlingen hun reflectieverslag uit hun duim omdat *'dat ook nog ff moet'* (Luken & Reynaert, 2009). Daarom is het goed om te kijken hoe in de ontwikkeling van kinderen en jonge mensen de rijping tot een zelfstandige reflectie (oordeelsvorming) in het algemeen verloopt en daarbij aansluiting te zoeken (Krijger, 2016). Uit deze opdracht is zoals gezegd naar voren gekomen dat het waardevol is om deze leerlingen (11^e klas) mee te laten spreken in hun eigen beoordeling. Bovendien is het voor het leerproces van de leerling beter als de leerling zelfkritisch leert te kijken naar het gemaakte product (Wiliam, 2016; Modderkolk, 2015).

Een goed doorleefde reflectie heeft een cyclisch karakter waarbij de leerling leert afstand te nemen van het eigen werk en vervolgens leerpunten formuleert om vanuit daar zijn leerproces weer verder ter hand te nemen (Wiliam, 2012). Het is van belang dat de leraar en leerling hierover in dialoog gaan met elkaar (Procee & Visscher-Voerman, 2004). Wanneer – in antwoord op de centrale vraag (pagina 1) – aan deze voorwaarden wordt voldaan zal een opdracht waarbij de leerling mee kan spreken in de beoordeling winnen aan kiemkracht en wordt het verhaal achter een cijfer duidelijk (Krijger, 2018). Dit versterkt het eigenaarschap van de leerling bij zijn eigen leerproces *en* zijn enthousiasme om te leren en te blijven leren.

Referenties

- Allen, D., & Tanner, K. (2006). Rubrics: tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE - Life Sciences Education*, (3), 197–203.
- Arter, J. A., & Chappuis, J. (2006). *Creating & recognizing Quality Rubrics*. Boston: Pearson.
- Biesta, G. (2012). *Goed onderwijs en de cultuur van het meten. Ethiek, politiek en democratie*. Den Haag: Boom Lemma.
- Biesta, G. (2015). *Het prachtige risico van onderwijs*. Culemborg: Phronese.
- Biesta, G. (2018). *De terugkeer van het lesgeven*. Culemborg: Phronese.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5, (1), 7-74.
- Black, P., Harrison, C., Lee, C. M., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi delta kappan*, 86, (1), 8-21.
- Bregman, R., & Frederik, J. (2016). *Waarom vuilnismannen meer verdienen dan bankiers*. Amsterdam: De Correspondent.
- Clandinin, D., & Connelly, F. (2000). *Narrative Inquiry. Experience and Story in Qualitative Research*. Francisco, USA: Jossey Bass. Wiley Imprint.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130-144.
- Koch, T. (2010). *Guidelines European Portfolio Certificate*. Brussel: EPC-group / ECSWE.
- Krijger, A. (2016). *Contemplatieve reflectie voor portfolio en actieonderzoek*. Antwerpen: Via Libra.
- Krijger, A. (2018). *Wat maakt een evaluatieproces kiemkrachtig?*. Leiden: Hogeschool Leiden.
- Luken, T., & Reynaert, W. (2009). *Puzzelstukjes voor een nieuw paradigma?* Eindhoven/Tilburg: Lectoraat Career Development Fontys Hogeschool HRM en Psychologie.
- Mayo, A. (2015). *Lectorale rede: Autonomie in verbondenheid*. Leiden: Hogeschool Leiden.
- Mayo, A. (2019). Brede vorming door het periodeonderwijs zichtbaar en inzichtelijk maken. In: *Opbrengsten van de bovenbouwconferenties*. Driebergen: Vereniging van vrijescholen.
- Modderkolk, W. (2015). *Essayopdracht: Economische filosofie*. Zeist: Stichtse Vrije School.
- Modderkolk, W. (2017). *Toetsing getoetst aan de doelstellingen van het vrijeschoolonderwijs*. Amsterdam: Hogeschool van Amsterdam.
- Mortelmans, D. (2007). *Handboek kwalitatieve onderzoeksmethoden*. Leuven/Voorburg: Acco.
- Polak, B. (2008, november 21). *Game Theory*. Opgehaald van Open Yale Courses: <https://oyc.yale.edu/economics/econ-159>
- Ponte, P. (2012). *Onderwijs en onderzoek van eigen makelijk. Onderzoek met en door leraren*. Den Haag: Boom Lemma.
- Procee, H., & Visscher-Voerman, I. (2004). Reflecteren in het onderwijs: een kleine systematiek. *VELON Tijdschrift voor lerarenopleiders*, 25, (3), 37-44.
- Rawson, M. (2010). Sustainable teacher learning in Waldorf education: A socio-cultural perspective. *Research on Steiner Education*, 1, (2), 26-42.
- Rawson, M. (2015). Assessment: A Waldorf perspective. *Research Bulletin Autumn/Winter 2015*. 20, (2), 30-42.
- Sadler, D. (1989). Formative Assessment and the Design of Instructional Systems. . *Instructional Science*, 18, (2), 119–44.
- Sedláček, T. (2012). *De economie van goed en kwaad*. Schiedam: Scriptum.
- Sluijsmans, D. (2019, januari 13). *Neem formatief niet te snel voor lief: tien lessen van Dylan Wiliam (5 t/m 7)*. Opgehaald van Toetsrevolutie: <http://toetsrevolutie.nl/?p=490>

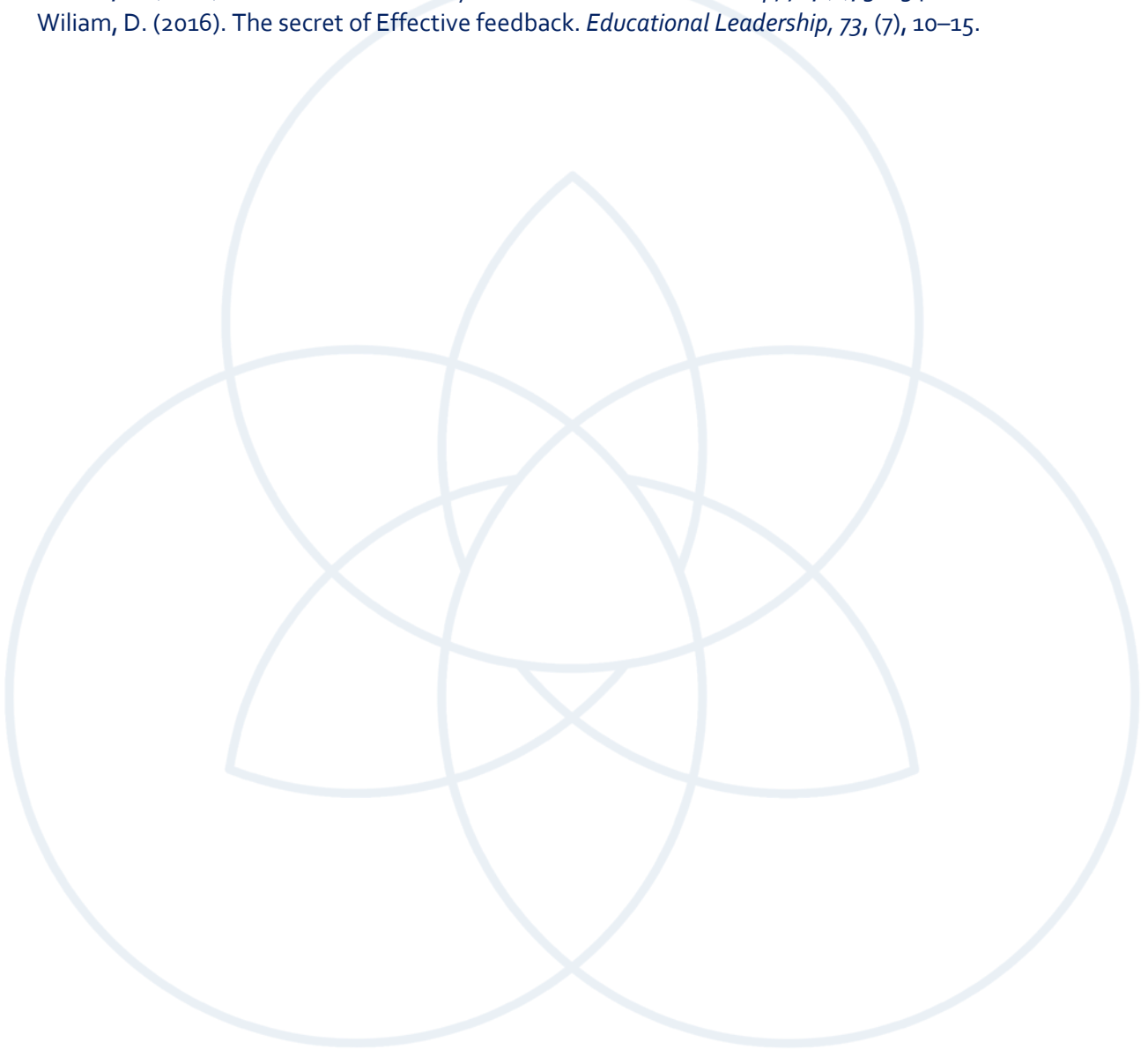
Steiner, R. (1996). *Geisteswissenschaftliche Behandlung sozialer und pädagogischer Fragen*. Dornach: Rudolf Steiner Verlag.

Steiner, R. (2004). *Opvoeding en onderwijs*. Zeist : Christofoor.

Straetmans, G. (2006). *Bekwaam beoordelen en beslissen*. Deventer: Saxion Hogescholen.

Wiliam, D. (2012). Feedback: Part of a System. *Educational Leadership*, 70, (1), 30–34.

Wiliam, D. (2016). The secret of Effective feedback. *Educational Leadership*, 73, (7), 10–15.





Over de auteurs

Ariëlla Krijger was tot 2018 als onderzoeker verbonden aan het lectoraat Waarde(n) van Vrijeschoolonderwijs bij de Hogeschool Leiden.

Wouter Modderkolk is leraar economie, decaan en projectleider op de Stichtse Vrije School in Zeist en is als onderzoeker verbonden aan het lectoraat Waarde(n) van Vrijeschoolonderwijs bij de Hogeschool Leiden.



Deze publicatie is beschikbaar onder de licentie Creative Commons Naamsvermelding-NietCommercieel-GelijkDelen 4.0. Wanneer je gebruik wilt maken van dit werk, hanteer dan de volgende methode van naamsvermelding: Ariëlla Krijger en Wouter Modderkolk, Achter elk cijfer schuilt een uniek verhaal. Praktijkvoorbeeld van een kiemkrachtig beoordelingsproces, CC-BY-NC-SA 4.0 gelicenseerd.

De volledige licentietekst is te lezen op <https://creativecommons.org/licenses/by-nc-sa/4.0/> of stuur een brief naar Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

