

Wanneer heeft iemand iets geleerd?

Dominique Sluijsmans en Desirée Joosten-ten Brinke

Eén van de meest complexe uitdagingen waarvoor leraren en lerarenopleiders in hun onderwijspraktijk staan is de vraag wanneer leerlingen c.q. studenten daadwerkelijk hebben geleerd. Om te kunnen vaststellen wanneer iemand heeft geleerd, is kennis over beoordelingsprocessen en aanpalende leertheorieën essentieel. Deze kennis kan het beste beschreven worden aan de hand van de volgende vijf vragen: 1) Waarom wordt beoordeeld? 2) Wat wordt beoordeeld? 3) Hoe wordt beoordeeld? 4) Wie beoordeelt? en 5) Wanneer wordt beoordeeld? Alvorens de vijf vragen worden beantwoord, zal het begrip beoordelen beknopt worden gedefinieerd en geplaatst in een historische context.

Wat is beoordelen en waar heeft het zijn oorsprong?

Hofstee (1999) definieert beoordelen in zijn algemeenheid als *‘het geval waarin mensen, op gezag van anderen, de kwaliteit(en) van iets van iemand vaststellen’* (p. 14). Het formele beoordelen in de context van leren kent een lange geschiedenis. Al in de 19^e eeuw was het de leraar die mondeling de voortgang van de lerende rapporteerde aan de ouders. Pas toen aan het einde van die eeuw de populatie van lerenden sterk toenam, begon men klassen te formeren op basis van leeftijd, het zogenaamde leerstofjaarklassensysteem. Er verschenen nieuwe methoden van beoordelen, zoals formele voortgangsrapportages. Leraren schreven op welke vaardigheden de lerenden al beheersten en welke nog ontwikkeld moesten worden. Deze vorm van beoordelen was de eerste vorm van een zogenaamd ‘narratief rapport’. Door de leerplichtwet rond de start van de twintigste eeuw groeide de lerende-populatie enorm, met als gevolg dat de onderwijshouden specifiek en de populatie van lerenden meer divers werden. Terwijl de lagere scholen nog de narratieve rapporten aanhielden om voortgang in leren te documenteren, werkten leraren in het voortgezet onderwijs al snel met percentages en cijfers. Deze manier van beoordelen die wij tot op de dag van vandaag nog gebruiken komt voort uit een industrieel onderwijsmodel, met kenmerken als klassikaal onderwijs, vakgericht lesgeven, aanbod- en methodegestuurde leerstof. Het heersende beoordelingsmodel betreft het *‘normgerichte beoordelen’*, waarbij lerenden worden afgezet tegen een gemiddelde van de groep. Deze vaak gestandaardiseerde beoordelingen veronderstellen uniforme lerenden en het bestaan van een ‘gemiddelde leerling/student’: het niveau van de groep is bepalend voor het individuele resultaat.

Waarom beoordelen?

Inmiddels laat onderzoek zien dat een industrieel onderwijsmodel, waarbij wordt uitgegaan van het bestaan van een gemiddelde student of leerling, de aangeboren nieuwsgierigheid en motivatie om te leren kan belemmeren en wellicht ontoereikend is in de 21^e eeuw (Bolocofsky & Mescher, 1984). Velen vinden dat het cijfersysteem te ver af staat van het echte leren en onderwijzen. Er komt meer aandacht voor een zogenaamd ‘organisch’ onderwijsmodel dat is gericht op het leerpotentieel van elke lerende (Boud, 2000). Kenmerken van dit model zijn differentiatie in werkvormen, leermiddelen en leerstof naar niveau, tempo, timing, belangstelling en een werkwijze die recht doet aan de unieke lerende (De Vries, 2007; Sluijsmans, 2008). Het sluit aan bij de ontwikkeling van zogenaamde assessments. Het woord ‘assessment’ is afgeleid van het Latijnse *‘assedere’* dat zoveel betekent als ‘naast iemand zitten’: samen met een lerende werken aan de ontwikkeling van kennis, vaardigheden en houdingen. De paradigmashift van een industriële naar een organische benadering van onderwijs vertaalt zich ook in de functies c.q. doelen van beoordelen, waarbij meer aandacht komt voor beoordelen in functie van leren. Daarbij wordt de balans gezocht tussen drie soorten functies:

- de *summative* functie: deze is gericht op het toetsen van een behaaiing over het niveau van de lereade met de bedoeling hierover aan anderen te rapporteren en/of verzamelde informatie te gebruiken voor de eigen behaaiingscyclus. Aan de hand van de verzamelde informatie worden afspraken gemaakt over de ontwikkeling van de lereade met het oog op de geseide leerdoelen dan wel over zijn/haar positie ten opzichte van een referentiegroep. Deze functie heeft vaak betrekking op beoordelingen over zaken of stages, of over toetsen of selecteren. Deze functie past bij het productiemodel van onderwijs.
- de *formative* functie: deze is gericht op het verzamelen van informatie over de lereade tijdens het leerproces. Deze informatie wordt verzameld, vastgelegd en geïnterpreteerd met de bedoeling deze te gebruiken in het onderwijsproces zelf. Bij formatief beoordelen zijn criteria en standaarden het uitgangspunt voor het verzamelen van informatie. In een review van 250 studies komen Black & William (1998) tot de conclusie dat toetsen die in functie van leren staan resulteren in significante en substantiële leeropbrengsten. Hoewel alle lerenden hiervan profiteren, is de leerwinst voor laag presterende lerenden groter dan voor andere lerenden.
- de *diagnostic* functie: deze is gericht op het zelf leren reguleren van het leerproces, waarbij voldoende ruimte is voor het zogenaamde 'spijfliciteitsovervloedige' (De Vries, 2007). Daarnaast beoordelen vereest een andere kijk op beoordelen omdat het vooral gericht is om leren in de toekomst te stimuleren dan juist vast te stellen of in het verleden en heden iets is gebeurd. In dit perspectief gaat duurzaam beoordelen bij het idee van een leven lang leren. Omdat kennisgebieden in hoog tempo veranderen en werkwijzen en technieken voortdurend veranderen, is het belangrijk lerenden te leren hoe zij voortdurend nieuwe kennis en vaardigheden eigen kunnen maken.

Wat beoordelen?

Voorts is voor elke vorm van beoordelen inzicht nodig wat en op welk niveau wordt beoordeeld. Wat beoordeeld moet worden is afhankelijk van de doelen van het onderwijs (kennis, vaardigheden, attitudes, competenties). Deze doelen worden afgeleid van de gewenste vaardigheden in een beroep of bepaald kennisdomein (bijvoorbeeld rekenen, taal of geschiedenis) en worden geoperationaliseerd in beoordelingscriteria en standaarden. Het vaststellen van beoordelingscriteria en standaarden gebeurt vooral door een groep van experts. Criteria kunnen betrekking hebben op de cognitieve, affectieve of motorische aspecten van de te leren kennis (kern of lereade, de 'beheersing'), vaardigheid (aan de lereade de kennis zoveel toepassen?), en/of op attitudeaspecten (houd de lereade zich aan de juiste regels en procedures?). Een overzicht van beoordelingscriteria vooreset in een gedetailleerde beschrijving van de gewenste leeruitkomsten.

Om vervolgens vast te stellen of een lereade iets heeft geleerd zijn goede beoordelingsmaten essentieel waarbij leren en beoordelen worden gekoppeld, omdat deze lerenden aanzetten tot het werken aan beoordelingscriteria die gericht zijn op behaaiing van kennis, vaardigheden en houdingen. Beoordelingsmaten kunnen de lereade in dit perspectief in vele vormen worden aangeboden zoals opdrachten, situaties, problemen of projecten. Om betrouwbaar vast te kunnen stellen wat de lereade heeft geleerd zijn vele, gevarieerde beoordelingsvormen nodig, afhankelijk van de kennis en vaardigheden en onderliggende criteria en standaarden die moeten worden gedemonstreerd (Shujumani, Straatman, & Van Merriënboer, 2009).

Hoe beoordelen?

Bij het beantwoorden van deze vraag hebben we te maken met drie aspecten: het beoordelingsmodel, het beoordelingsinstrument en het resultaat van de beoordeling en de feedback.

Er zijn grofweg drie beoordelingsmodellen te onderscheiden. Aan elk beoordelingsmodel liggen andere criteria en standaarden ten grondslag. Het eerste model heeft het *normatieve Assesdén*, waarbij leerlingen worden afgezet tegen een gemiddelde van de groep. Dit resulteert in een zogenaamde *Gaussiane* (vernoemd naar de Duitse wiskundige Friedrich Gauss). Deze vaak gestandaardiseerde beoordelingen verspreiden zich over uniforme leerlingen en het bestaan van een 'gemiddelde leerling'. Het niveau van de scoregroep is bepalend voor het individuele resultaat. Een tweede beoordelingsmodel is het *crisis/impulsief Assesdén*, een model dat we zelden aantreffen in instrumenten. Voordelig hangt niet in het middelbaar en hoger beroeps onderwijs en gebaseerd is op competenties en kwalificaties. Leraren/leraresopleiders vergelijken leerlingen in deze benadering niet zover met de norm van de groep, maar beoordelen hen op een aantal beoordelingscriteria en bijbehorende standaarden. Bij beoordelen gericht op duurzaam leren waarbij het bevorderen van betrokkenheid van leerlingen centraal staat past een *positief beoordelingsmodel*. In dit model worden beoordelingen van leerlingen niet zover vergeleken met het gemiddelde van hun klas of met vooraf vastgestelde criteria en standaarden, maar worden individuele beoordelingen die zijn uitgesplitst aan het niveau van de leraar vergeleken in de tijd. Er is sprake van duurzaam beoordelen omdat criteria en standaarden flexibel kunnen worden aangepast en de beoordelingen zich niet meer beperken tot een momentopname. Onderzoek van Slavin (1980) laat zien dat een positief beoordelingsmodel tot beter leren leidt dan een normgericht beoordelingsmodel.

Om te beoordelen zijn verschillende instrumenten beschikbaar. De keuze voor een instrument wordt ingegeven door het doel en de visie die de leraar/leraresopleider heeft met betrekking tot de beoordeling en praktische randvoorwaarden zoals organisatievaardigheid en tijd. Soms zijn er ook beperkende kenmerken met verschillende vraagvormen, mondelinge toetsen, performance assessments, peer assessments, self assessments, of portfolio assessments zijn veel voorkomende instrumenten. Het resultaat van beoordelen kan ook variëren. Beoordelingen kunnen gepresenteerd worden in de vorm van cijfers, uitgangspunten, bewijzen van de leraar, ingevulde beoordelingschema's ('rubrics'), ontwikkelingsplannen of mondelinge besprekingen. Voor degene die wordt beoordeeld moet duidelijk zijn wat de betekenis van de beoordeling is.

Naast de keuze van een geschikte beoordelingsvorm, is in elke vorm van beoordelen de feedback de leraar van leraaropleider) van belang op basis van hun positieve en cruciale aspect. Feedback is het instrument waarmee de leraar/leraresopleider verzamelde informatie over het leren van leerlingen omvat maar zelden afbreken om het leren te verbeteren (Sadler, 1989). Deze leerbevorderende feedback is gericht op het veranderen van de verschillende manieren een taak te begrijpen en/of de manier van de leraar en een gevormd resultaat (Sajjama, Weijnen, Castelijn & van de Ven, 2011). Beoordelen zonder het bieden van effectieve feedback heeft het leren en de zelfregulatie van leerlingen niet verder (Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006). Dit onderzoek wijst dat voor de laag presterende leerling bij gedifferentieerde, specifieke en geïndividualiseerde manieren van feedback op taken (Owen, 1983). Onderzoek toont tevens aan dat de feedbackvaardigheid van de leraar één van de belangrijkste factoren is voor succesvol onderwijs (Hattie, 2009). Nu sommigen leraar/leraresopleiders nog vaak alleen het resultaat van een beoordeling ('*Knowledge of Results*'), voldoende, cijfer, geslaagd, goed gedaan), waardoor de leraar geen aandacht wordt geboden voor verdere ontwikkeling.

Feedback is gericht op het beantwoorden van de volgende drie vragen:

- Waar gaat de lerende naar toe (wat zijn de geselecteerde taken)? Dit kan worden benoemd als de *Feedup*.
- Hoe heeft de lerende de taak uitgevoerd (welke vooruitgang wordt geboekt)? Dit kan worden benoemd als de *Feedback*.
- Hoe gaat de lerende verder terwijl aanpak is nodig om tot groei te komen? Dit kan worden benoemd als de *Feedforward*.

Er zijn vervolgens vier feedbackniveaus te onderscheiden. Feedback op *naar/voorwaarts* is effectief wanneer sprake is van misinterpretatie van de taak, en dient om de lerende aanwijzingen te geven om tot een juiste interpretatie van de opdracht te komen. Feedback op *pro/voorwaarts* is effectief wanneer lerenden op zoek zijn naar een strategie om een bepaalde taak uit te voeren. Feedback op het *afwaarts* van *zelfopvallende* onderdelen de lerende bij het vasthouden van de aandacht en inzet bij de uitvoering van de taak, bij het laten attribueren van succes en de positieve feedback van anderen en daarmee bij het vergroten van het zelfvertrouwen. Feedback die betrekking heeft op het *persoonlijk afwaarts* (complimenteren, prijzen) kan leiden tot het vermindern van lastige taken en het minimaliseren van inzet. Hoeveel deze vorm van feedback waardevol kan zijn 'in-het-moment' toont onderzoek aan dat dit niet tot verbeterde leerresultaten leidt (zie ook Black & William, 1998).

Wie beoordelen?

In het beoordelingsproces kunnen verschillende personen betrokken zijn. Welke beoordelaars worden ingezet is afhankelijk van de benodigde dekkingheid. In formele onderwijsinstellingen is de docent wettelijk verantwoordelijk voor het beoordelen van de student. Ter beoordeling van de betrouwbaarheid van beoordelingen is het noodzakelijk om meerdere beoordelaars in te zetten. Het wettelijk spel bij competentiegerichte beoordelen een belangrijke rol om de prestaties van de lerende vanuit de praktijkvisie te beoordelen. De beoordelaar kan hierbij een lid van de lerende in het bereik zijn, maar ook klanten of partners zijn mogelijk inbaar als beoordelaar. Omdat de lerende echter de enige is die zichzelf 'meesteem' in het leertraject en zichzelf voortdurend kan 'zien' en ook de kennishorizonten en omgevragen zo wisselend zijn, zal hij in beoordelende mate in staat moeten zijn leren te reguleren. Dit betekent dat hij een grotere betrokkenheid krijgt in het beoordelingsproces en het beoordelen samen met een leeropvoeder vormgeeft. In de historie van het beoordelen is de lerende lange tijd niet als voorwaardige partner beschouwd. Het beoordelen was vooral een zaak van partijen buiten de lerende. Opedagogische vormen van beoordelen behelmenen kennis onder het beoordelen niet toebehoort aan de lerende, maar iets is dat de lerende ondergaat (Boud, 2000). Onderzoek laat zien dat beoordelingen die unilateraal gericht zijn op certificering en selectie negatief zijn voor leren, omdat de lerenden behelmenen om verantwoordelijkheid te nemen zich te motiveren en te reguleren (Hattie & Timperley, 2007; Ryan & Deci, 2000). Beoordelen op deze wijze leidt tot 'aangestuurde afhankelijkheid' (Voorke, 2001). De laatste jaren is in wetenschappelijke publicaties sprake van een groeiende belangstelling voor de actieve rol van de lerende in het onderwijsproces door het inzetten van vormen van zelf- en peerbeoordeling (Shijmans, Dochy, & Moerkerke, 1999). Het geven van commentaar lerenden over de wijze waarop ze worden beoordeeld, wanneer ze worden beoordeeld en door wie ze worden beoordeeld, kan leiden tot een activerende betrokkenheid en verantwoordelijkheid van lerenden. Dit kan vervolgens tot een hogere motivatie en tot meer kennis leiden dan wanneer het beoordelen extern wordt aangestuurd (Wulf, Raupach, & Pfeiffer, 2005).

Wanneer beoordelen?

Uit de voorgaande antwoorden op de vragen blijkt dat beoordelen – afhankelijk van de functie (formatief, summatief of diagnostisch) op verschillende momenten in het leerproces kan worden ingezet. Vanuit een summatief perspectief is beoordelen zinnig aan het begin van een leerpad (Is de leraar geschikt voor de opleiding? Wat is de voorkennis?), tijdens het leerproces (Hoeveel de leraar voldoende voortgang op de bepaalde criteria en standaarden? Is de leraar klaar voor een volgend niveau?) en aan het einde van een leertraject (Wat wordt de vervolgopleiding? Is de leraar klaar voor het beroep?). Formatieve beoordelingen zijn vooral tijdens het leerproces van belang, omdat deze de leraar informatie bieden over zijn ontwikkelings- en de volgende te nemen stappen. Bij duurzaam beoordelen vallen leren en beoordelen als het ware samen en is sprake van een continue beoordelings- e.g. leerproces. Bij duurzaam beoordelen wordt in alle fasen van het leerproces beoordelingsinformatie verzameld die de leraar vooral ondersteunt in het zelf leren reguleren van zijn leren.

Van groot belang is dat op het niveau van de school en de lerarenopleiding een mix van beoordelingsvormen aan bod komt, waarmee de kwaliteit van het toetsprogramma wordt geborgd. Vragen die daarbij gesteld kunnen worden zijn: Is er voldoende informatie verzameld over de leraar op basis waarvan op een betrouwbare en valide wijze kan worden gezegd dat er is geleerd? Hoeveel de leraar voldoende leren zien dat bij of zij ook nu het verlaten van de school of opleiding het leerproces kan reguleren? Passen de beoordelingsvormen van de lerarenopleiding bij de toekomstige beroepspraktijk? Zijn leraren(opleiders) voldoende onderlegd in kennis en kunde over beoordelen? In dit opzicht heeft Baartman en andere (2006) een waardevol kader waarmee de kwaliteit van beoordelen op programmatief niveau kan worden gemonitord en vanuit die analyse te komen tot een verantwoord beoordelingsmodel binnen de school en lerarenopleiding.

Tot slot

Om goed te kunnen vaststellen of er is geleerd, is niet alleen een nauwe aansluiting tussen leren en beoordelen noodzakelijk (Biggs, 1996), maar vooral een heldere visie op het doel van beoordelen. Aangezien beoordelen al lang niet meer uitsluitend de functie heeft van selectie en certificering, maar steeds meer is gericht op het versterken van (lange termijn) leren kan en mag de theorie over beoordelen niet meer los staan van andere theorieën over (ontwikkelen van) leerprocessen bij studenten en lerenden. Veel theorie in de kennisbasis van lerarenopleiders is dan ook van toepassing op de wijze van beoordelen. Hoe ga ik om met heterogeniteit in beoordelen? Hoe bepalen ik lerenden in het beoordelingsproces? Hoe vindt beoordelen in de beroepspraktijk plaats? Hoe kan ik het samenwerkend leren binnen beoordelen vormgeven? Hoe bepalen ik de kwaliteit van beoordelen? De kennisbasis over beoordelen is in dit opzicht interdisciplinair en nauwelijks te begrenzen.

Literatuur

- Baartman, L. K. J., Baartmans, Th. J., Kirschner, P. A., & Van der Vlieten, C. P. M. (2006). The wheel of competency assessment: Presenting quality criteria for competency assessment programs. *Studies in Educational Evaluation, 32*, 153-177.
- Biggs, J. (1996). *Enhancing teaching through constructive alignment*. *Higher Education, 32*, 347-364.
- Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*, 7-74.
- Bokorok, D. N., & Mescher, S. (1984). Student characteristics: Using student characteristics to develop effective grading practices. *The Diverse Teacher, 6*, 11-23.
- Boud, D. (2001). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*, 151-167.

De Vries, B. (2007). *Vrijheid, Mijheid? Over vaaggestemd leren op flexibele scholen*. Nijmegen: Hogeschool Arnhem en Nijmegen.

Hattie, J. (2008). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement*. Taylor & Francis.

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

Hobson, W. R. B. (1999). *Principles van beoordeling*. Lisse: Swets & Zeitlinger B. V.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31, 199-218.

Owen, D. E. (1983). Evaluation systems and attributional tendencies in the classroom: A sociological approach. *Journal of Educational Research*, 76, 307-312.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68-78.

Sadler, R. D. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 139-144.

Slavin, R. E. (1980). Effects of individual learning expectations on student achievement. *Journal of Educational Psychology*, 72, 528-524.

Sluijtmans, D. M. A. (2008). *Beoordelen bij hogeschoolen (lecturale rede)*. Nijmegen: Hogeschool van Arnhem en Nijmegen.

Sluijtmans, D. M. A., Doehly, F., & Moerkelke, G. (1999). Creating a learning environment by using self, peer- and co-assessment. *Learning Environments Research*, 1, 283-319.

Sluijtmans, D. M. A., Stehouwer, G., & Van Merriënboer, J. (2008). Integrating authentic assessment with competency based learning: The protocol portfolio scoring. *Journal of Vocational Education and Training*, 19, 9-12.

Sluijtmans, D. M. A., Weijnen, S., Castejns, J., & van de Ven, H. (2011). Leerkolofnback als motor voor zelfreguleerd leren. In J. Castejns, M. Segers, & K. Struyven (Eds.), *Evalueren om toe te leren*. Bussum: Coutinho.

Wolf, G., Knappich, M., & Pfeiffer, F. (2005). Self-controlled observational practice enhances learning. *Research Quarterly for Exercise and Sport*, 76, 107-111.

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