

„Mastery learning” voor de hogere cognitieve doelstellingen*

A. E. N. ROMMES**

Centrum voor Didactiek en Onderzoek van Onderwijs T.H. Twente

Samenvatting.

Bij de realisering van op mastery gericht onderwijs zijn de volgende vragen essentieel: wat moet men verstaan onder beheersing van de stof en hoe moet die getoetst worden.

De aard van de onderwijsdoelstellingen bepaalt het criterium wanneer een student wèl en wanneer hij nièt mastery heeft bereikt.

Het onderscheid tussen doelstellingen op het nivo van kennisreproductie en de hogere cognitieve doelstellingen heeft vergaande consequenties. Dit geldt onder meer voor vraagstukken t.a.v. de ideale moeilijkheidsgraad van mastery toetsen, de relatie met intelligentie, de betrouwbaarheid en de methodes voor het bepalen van de cesuur tussen voldoende en onvoldoende.

1. Inleiding.

Mastery learning wordt vaak gepresenteerd als een nieuwe visie op onderwijs. Bij mastery learning streeft men er naar het onderwijs zo in te richten dat bijna elke student tot beheersing van de leerstof komt. Dit kan o.a. bereikt worden door het opdelen van de stof in leertaken, ieder afgesloten door een toets. De student mag pas

* Als Nederlands equivalent voor de term 'Mastery learning' komt het woord 'beheersingsleren' in aanmerking.

** De opmerkingen van medewerkers van het CDO, van de T.H. Twente en van het Centrum voor Onderzoek van het Wetenschappelijk Onderwijs, Universiteit van Amsterdam hebben tot verbetering en verduidelijking bijgedragen. De verantwoordelijkheid voor eventuele onjuistheden blijft uiteraard voor rekening van de auteur.

doorgaan naar de volgende taak wanneer de vorige beheerst wordt. De tijdsduur van een mastery cursus is flexibel en hangt af van de voortgang van de individuele student. Het begrip mastery learning is afkomstig van Morrison (1926). Bloom (1971) en Carroll (1971) hebben de mastery visie op grotere schaal ingang doen vinden. Het septembernummer van het tijdschrift *Onderzoek van Onderwijs* geeft een duidelijke beschrijving van het mastery learning model (Kramers-Pals e.a. 1973). In hetzelfde nummer is een kritisch artikel over het mastery learning model opgenomen, onder de titel 'Ieder kan op den duur alles leren; gaat u ook nooit iets boven de pet?' (De Groot, 1973).

Een aantal kenmerken van mastery onderwijs kan de grote populariteit ervan verklaren. Gedacht kan bijvoorbeeld worden aan het adaptieve karakter, in tegenstelling tot het selectieve karakter van 'klassiek' onderwijs. Ook de nadruk op expliciete doelstellingen, gekombineerd met absolute normen ligt goed in de onderwijsmarkt. Langzamerhand is in samenhang met de mastery learning visie een aantal ideeën in omloop geraakt die onjuist zijn, of die ten onrechte aan het mastery model toegeschreven worden. Een voorbeeld van zo'n onjuiste opvatting is de idee dat iedereen alles kan leren, wanneer er maar voldoende tijd beschikbaar is. Deze ekstreme opvatting wordt door De Groot weerlegd in het eerder genoemde artikel. Hieronder zal ingegaan worden op een aantal andere opvattingen die men bij het realiseren van mastery onderwijs vaak impliciet als uitgangspunten hanteert. Bij een explicitering van deze opvattingen zal blijken dat ze maar een beperkte geldigheid hebben, of dat ze minder essentieel zijn voor mastery learning dan wel

gesuggereerd wordt. Onder meer zullen besproken worden: de opvatting dat bij mastery onderwijs alle doelstellingen beschreven moeten worden; dat volledige beheersing van alle doelstellingen betekent dat items van mastery toetsen door vrijwel iedere student goed gemaakt moeten worden; dat de klassieke betrouwbaarheidstheorie niet geldig is voor mastery toetsen; dat bij mastery toetsen absolute (criterion referenced) normen gehanteerd moeten worden. Voordat we op deze opvattingen ingaan beschrijven we twee trends die aan mastery onderwijs ten grondslag liggen. Verder zal blijken dat het noodzakelijk is nader te konkretiseren welke onderwijsdoelstellingen met mastery learning bereikt moeten worden. Deze konkretisering wordt gegeven in paragraaf 3.

2. Mastery learning.

De omslag van het boekje 'Mastery learning, theory and practice' (Block, ed., 1971) bevat een figuur dat de mastery visie in grafische vorm verbeeldt. Dit figuur wordt wel het grafisch symbool voor de mastery visie genoemd (Warries, 1973).

De gedachte die dit figuur overbrengt is duidelijk. Het bevat twee verdelingen van studieresultaten. Op de niet getekende vertikale as is de frekwentie, het aantal personen, afgezet. Op de horizontale as de studieresultaten. De 'normale' symmetrische verdeling zou betrekking hebben

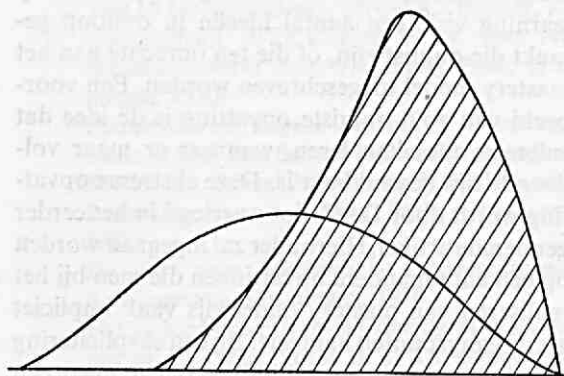


Fig. 1. Studieresultaten in mastery onderwijs en 'klassiek' onderwijs.

op studieresultaten bij 'klassiek' onderwijs: er zijn grote verschillen tussen de resultaten van de verschillende studenten, naast zeer goede zijn er even veel zeer slechte resultaten. De scheve verdeling* zou gevonden worden bij mastery onderwijs: de verschillen zijn kleiner geworden, de zeer slechte resultaten zijn verbeterd.

Figuur 1 laat niet zien hoeveel studenten in mastery onderwijs slagen. Wel is uit de mastery visie duidelijk dat het de bedoeling is dat alle of vrijwel alle studenten slagen, terwijl de voorwaarde voor slagen in mastery onderwijs is dat de student de stof beheerst.

Deze twee laatste punten verklaren voor een belangrijk deel de snel groeiende populariteit van de mastery visie. Ze sluiten aan bij twee actuele trends in het onderwijskundig denken; namelijk de opvatting dat het mogelijk is onderwijs aan te passen aan de capaciteiten van de studenten (adaptief onderwijs) en de opvatting dat onderwijs tot doel heeft de student tot volledige beheersing van de stof te brengen. Op deze beide trends zal nu nader ingegaan worden.

2.1. Adaptief onderwijs.

In de lijn van de eerste trend kan de mastery visie gezien worden als een reactie op het selectieve karakter van veel onderwijs. Wanneer studenten onvoldoende van onderwijs profiteren wordt dit toegeschreven aan hun ongeschiktheid voor dat onderwijs, en de daarbij passende actie is afval of selectie. De gedachte dat niet de studenten aangepast moeten worden maar dat ook van het onderwijs aanpassing gevraagd kan worden vindt weerklink in de mastery visie. Natuurlijk zijn er verschillen tussen studenten in de mate van geschiktheid (de een leert sneller dan de ander), maar deze verschillen kunnen opgevangen worden door gebruik te maken van onderwijsmethoden waarin variaties in tempo en totale tijd toegestaan zijn. Deze reactie op het selectieve onderwijs benadrukt dat *iedere* student succesvol

* De figuur bij Block bevat een onjuistheid die hier ter wille van de authenticiteit overgenomen is. De oppervlakte onder de twee curves moet namelijk gelijk zijn, dat is hier echter niet het geval.

aan onderwijs kan deelnemen. De resultaten van onderwijs volgens de mastery gedachte horen in deze lijn dan ook tot uiting te komen in slagen van 100%, of nagenoeg 100% van het aantal studenten dat deelneemt aan de mastery cursus.

2.2. *Beheersing van de stof.*

De andere onderwijskundige trend bij de mastery visie heeft te maken met de centrale positie die men de onderwijsdoelstellingen in het onderwijs toekent. Men is zich duidelijk bewust geworden dat men bij de konstruktie van onderwijs aan dient te geven wat met dat onderwijs bereikt moet worden. Zo moet men expliciteren welke vaardigheden de student moet aanleren, welke kennis verworven moet worden. Bij cursus-konstruktie begint men daarom met het formuleren van de doelstellingen bv. in de terminologie van Bloom, Gagné of Mager.

Voorbeeld van een doelstelling volgens Mager: de leerling moet binnen een tijd van 30 minuten tenminste 7 eenvoudige lineaire vergelijkingen korrekt kunnen oplossen.

Bij het formuleren van de onderwijsdoelstellingen beperkt men zich uiteraard tot de essentiële en belangrijke doelstellingen. In de mastery visie vindt men deze terug wanneer beklemtoond wordt dat een student tot beheersing van de stof moet komen.

2.3. *Het criterium voor beheersing.*

Het al dan niet beheersen van de stof vormt het criterium voor het al dan niet slagen voor het betreffende onderwijs. Studenten die de stof beheersen slagen, studenten die de stof niet beheersen zakken. Bij de toetsing van de resultaten van mastery onderwijs zou daarom verwacht moeten worden dat een twee-deling ontstaat. De verdeling van de resultaten in figuur 1 zou dan tweetoppig, dichotoom zijn. Dit is niet het geval. Kennelijk is het niet zo dat er een criterium is dat duidelijk diskrimineert tussen wel of niet beheersen.

Deze bevinding is op het eerste gezicht merkwaardig. Men zou verwachten dat, wanneer men uitgaat van een beschrijving van de doelstel-

lingen, daarmee ook vast te leggen moet zijn of de doelstellingen al dan niet bereikt zijn. De redenering die daarbij wel gekozen wordt is de volgende. Wanneer de onderwijsdoelstellingen expliciet geformuleerd zijn kan nagegaan worden of de prestaties van de studenten daarmee overeenstemmen. De items in de eindtoets worden zo gekozen dat ze een representatie vormen van de doelstellingen. Aangezien de doelstellingen bereikt moeten zijn eist men dat deze items goed gemaakt worden. Bij de toetsuitslag zou dat betekenen dat de student die de stof volledig beheerst een maximale score op de toets bereikt d.w.z. dat hij alle items goed heeft. Een toets-technische konsekventie is dan dat de items een P-waarde (moeilijkheidsgraad) van 100% hebben. Soms wordt een P-waarde van 100% of dicht daarbij dan ook als ideaal gezien voor mastery onderwijs. In het volgende hoofdstuk wordt nagegaan wat de waarde van deze redenering is, wanneer hij wel en wanneer hij niet opgaat.

3. *Doelstellingen.*

Om duidelijker zicht te krijgen op de criteria voor mastery is het nodig verder in te gaan op de vraag wanneer er sprake is van volledige beheersing van de doelstellingen, terwijl tevens de relatie tussen de doelstellingen en de methode om het te toetsen bekeken moet worden. Hierbij wordt uitgegaan van de klassifikatie van onderwijsdoelstellingstaken van Merrill en Boutwell (1973).

Merrill en Boutwell geven een klassifikatie van onderwijsdoelstellingen voortbouwend op het werk van Gagné en Bloom. Ze onderscheiden daarbij twee dimensies nl. een inhoudelijke en een gedragsdimensie. Langs de inhoudelijke dimensie wordt de leerstof beschreven; langs de gedragsdimensie de activiteit van de student. De inhoud van doelstellingen is te klassificeren door aan te geven of het gaat om een probleem, een principe (regel), een begrip (concept) of een gepaarde associatie. De gedragsdimensie van doelstellingen heeft betrekking op dat wat de student moet kunnen doen aan het eind van het

BEHAVIOUR	RULE FINDING			PROBLEM SOLVING (higher order rule using)	
	RULE USING		RULE USING	COMPONENT RULE USING	
	CLASSIFICATION		INSTANCE CLASSIFICATION	PRINCIPLE CLASSIFICATION	
	DISCRIMINATED RECALL	DISCRETE MEMORY and/or COMPONENT RECALL	DEFINITION and/or INSTANCE RECALL	RULE and/or SOLUTION RECALL	HIGHER RULE and/or SOLUTION RECALL
		PAIRED ASSOCIATE	CONCEPT	PRINCIPLE	PROBLEM
		CONTENT			

Fig. 2. Twee-wegklassifikatie van onderwijsdoelstellingstaken (volgens Merrill en Boutwell).

onderwijs. De volgende vier categorieën worden daarbij onderscheiden: kennisreproductie (discriminated recall), klassificeren, regels toepassen, regels ontdekken (zie fig. 2). We zullen nu iets nader ingaan op de gedragsdimensie van de doelstellingen en de aard van de beheersing die een student in de verschillende klassen kan bereiken.

a. *Reproductie op kennisniveau* (discriminated recall). Het gedrag van een student bestaat uit reproductie op kennisniveau, wanneer hij bij aanbieding van een prikkel bijvoorbeeld een symbool, een object of een gebeurtenis onmiddellijk antwoordt met een bijpassend symbool, object of gebeurtenis. De reproductie gebeurt uit het

geheugen, de student moet de combinatie van de prikkel en het bijbehorende antwoord gememoreerd hebben. Voorbeelden van dit soort reproductie zijn het benoemen van verschillende merken auto's, het reproduceren van vertalingen van vreemde woorden, het herkennen van een toonhoogte, het geven van een definitie van een begrip, het formuleren van een geleerd principe (regel) of een algoritme. Een voorbeeld van een doelstelling uit Mager op dit gedragsnivo:

‘In staat zijn alle bedieningsknoppen van de radarapparatuur korrekt te benoemen’, of:

‘reproduceren van een oplossingsschema van een bepaald type probleem.’

Bij onderwijsdoelstellingen die in deze categorie vallen blijkt de beheersing van de student wanneer hij bij aanbieding van de prikkel het bijbehorende antwoord reproduceert. De doelstellingen kunnen uitputtend beschreven worden door aan te geven welke prikkel – antwoord (S-R) associaties gekend moeten worden.

Op het nivo van kennisreproductie kunnen de onderwijsdoelstellingen van te voren uitputtend beschreven worden. Het gaat om een gesloten, eindig geheel van kenniseenheden, die ook alleen maar te onderwijzen zijn door ze allemaal in het onderwijs op te nemen. Deze eindigheid bij kennisdoelstellingen vormt een tegenstelling met de andere klassen van doelstellingen in het schema van Merrill en Boutwell. Bij de andere klassen is steeds sprake van mogelijkheden tot generalisatie of transfer van het geleerde. Meestal is de generalisatie of transfer zelfs de essentie van deze doelstellingen. De doelstellingen vormen een aanduiding van een niet-eindige populatie van gedragingen.

Laten we kort kijken naar de drie andere gedragsnivo's in de klassifikatie van doelstellingen: classification, rule using and rule finding.

b. *Classification*. Klassificeren is gedrag dat het meest geëigend is bij het leren van begrippen (concepts). Een student beheerst een begrip wanneer hij kan vaststellen of een voorbeeld al dan niet deel uitmaakt van het begrip, of dat hij

de eigenschappen van een begrip kan noemen, vooropgesteld dat deze nog niet eerder genoemd zijn. Bijvoorbeeld: de hond is een voorbeeld van een gewerveld dier. Het is niet mogelijk en ook niet praktisch een student alle voorbeelden van een begrip te laten geven aangezien de meeste begrippen in de praktijk een oneindig aantal voorbeelden bevatten.

c. *Rule using* (toepassen van regels). Het toepassen van regels vindt vooral plaats bij principes. Het gaat dan om het relateren van twee begrippen. Een student kan de regel toepassen wanneer hij zelf bij een gegeven voorbeeld van het ene begrip het bijbehorende voorbeeld van het andere begrip kan vinden. Bijvoorbeeld: stroomsterkte is omgekeerd evenredig met de weerstand.

d. *Rule finding*: higher order rule using. Het vinden van regels is van toepassing op problem solving situaties; de student moet bepalen welke speciale regel op het probleem van toepassing is. Het gaat meestal om een reeks van achtereenvolgens toe te passen regels.

4. *Konsekwenties voor mastery toetsen.*

De gevolgen voor de mastery opvatting lijken duidelijk. Alleen in het geval van doelstellingen op het nivo van kennisreproductie is het mogelijk volledig aan te geven wat de student wel en wat de student niet hoeft te beheersen. Bij de hogere cognitieve doelstellingen (classification, rule using, rule finding) is sprake van generalisatie of transfer. Dit is ook het voornaamste doel van het meeste onderwijs. De noodzakelijkerwijs beperkte leerstof wordt aangeboden met het doel bruikbaar te zijn voor andere situaties dan de behandelde. De nadruk die in het onderwijs op 'inzicht' gelegd wordt is vooral om de toepasbaarheid op nieuwe situaties te vergroten.

De betekenis voor de mastery toetsen kan bekeken worden aan de hand van de moeilijkheidsgraad (P-waarde). Bij items die geschreven worden voor doelstellingen in de klasse van kennis-

reproductie kan volledige beheersing geëist worden. Het zou onzinnig zijn eerst bv. door middel van een taakanalyse te bepalen welke kennis noodzakelijk is en later genoeg te nemen met minder dan volledige beheersing daarvan. Zou minder toegestaan zijn dan hoort die kennis bij voorbaat geschrapt te worden uit de leerstof. Dit betekent dat bij kennis-items in mastery onderwijs gestreefd kan worden naar P-waarden van 100%.

De moeilijkheidsgraad van items die betrekking hebben op hogere cognitieve doelstellingen zal heel anders liggen. Bij het klassificeren van een voorbeeld onder een begrip hangt het gemak waarmee dat gebeurt af van de duidelijkheid waarmee het voorbeeld tot het begrip behoort. Een voorbeeld dat veel concurrerende associaties oproept zal niet gauw tot het begrip gerekend worden. Ook bij rule-using en rule-finding zullen de in de leerstof behandelde regels of ketens van regels niet allemaal even automatisch en vanzelfsprekend toegepast kunnen worden. Veel items zullen door de student moeilijk geassocieerd worden met de behandelde stof. Het is echter de bedoeling verder te komen dan alleen de behandelde stof, daarom zullen deze items wel gevraagd worden. Dat de P-waarden van deze verder weg liggende items lager dan 100% liggen zal niemand verbazen*.

4.1. *De relatie met intelligentie.*

Aanhangers van de mastery visie zullen bij P-waarden lager dan 100% tegenwerpen dat de kans groot is dat deze items niet voldoen aan de eis van specificiteit voor de onderwezen stof. Er moet namelijk een goed onderscheid gemaakt worden tussen het meten van kennis en inzicht verworven in het onderwijs en het meten van cognitieve factoren die de student al eerder heeft. Bij moeilijke items bestaat het gevaar dat ze voornamelijk een beroep op intelligentie doen.

* In deze lijn ligt overigens ook de bevinding dat zowel docenten als studenten geneigd zijn moeilijke items begripsvragen te noemen (Vonk 1967, aangehaald in Hofstee 1973).

Toetsen moeten echter 'intelligentiezwak' zijn (Van Naerssen, 1973).

Bekijken we wat de student moet presteren bij items die een beroep op de hogere cognitieve vaardigheden doen, dan doet dit sterk denken aan een omschrijving van intelligentie als het vermogen *nieuwe* problemen op te lossen. De transfer van het bekende naar het nieuwe is nodig. Het ligt dan ook voor de hand dat de beantwoording van de items afhangt van de intellectuele capaciteiten.

De eis dat de items zo gekonstrueerd moeten zijn dat de beantwoording ervan afhangt van de kennis van de stof is als volgt empirisch te controleren. We vergelijken de prestaties op een item van een groep studenten die de stof niet bestudeerd heeft met een groep die de stof wel bestudeerd heeft (zie fig. 3).

In kwadrant I staan studenten die de stof bestudeerden en het item goed hebben. In kwadrant II de studenten die de stof nog niet bestudeerden en het item goed hebben. In kwadrant III staan de studenten die de stof niet bestudeerden

		niet bestudeerd	wel bestudeerd
item goed	II	I	
item fout	III	IV	

Fig. 3. Correlatie tussen bestudering en prestatie op het item.

en het item fout hebben. In kwadrant IV staan de studenten die de stof wel bestudeerden en het item fout hebben.

Wanneer de prestaties van de studenten afhangen van de bestudering van de stof en van intellectuele capaciteiten kunnen de prestaties als volgt over de kwadranten verdeeld zijn. De kwadranten I, III en IV zijn gevuld. Vulling van kwadrant IV is toegestaan: het item kan fout gemaakt worden omdat het een beroep doet op niet behandelde intellectuele vaardigheden. Kwadrant II moet leeg zijn, anders is er sprake van een item dat opgelost kan worden zonder kennis van de stof. Bij een dergelijke verdeling* over de kwadranten is kennis van de stof een voorwaarde voor het juist beantwoorden, het is echter niet de enige voorwaarde.

Het zou heel vervelend zijn wanneer het optreden van transfer van leerstof naar nieuwe situaties uitsluitend afhankelijk van intelligentie was. De effecten van onderwijs zouden dan wel erg beperkt blijven. Het zou beter zijn te spreken van een wisselwerking tussen de aanwezige capaciteiten en de nieuw aan te leren mogelijkheden. Een toename van de intellectuele capaciteiten naarmate meer onderwijs genoten werd, ligt daarbij voor de hand. Dit ligt ook in de lijn van de beleving van ieder die studeert; de belangrijkste activiteiten zijn daarbij: het zich eigen maken van de stof o.a. door het bedenken van nieuwe toepassingsmogelijkheden, nieuwe voorbeelden en nieuwe verbanden.

4.2. *Betrouwbaarheid van mastery toetsen.*

Het toepassen van de klassieke testtheorie op vorderingentoetsen wordt vaak als problematisch ervaren. Dit komt met name door de vraag of het bij een vorderingentoets, net als bij tests, gaat om het meten van één faktor. Bij het klassieke betrouwbaarheidsmodel wordt uitgegaan van het meten van één faktor, één dimensie. Wanneer het bij een vorderingentoets om enkele

onafhankelijke factoren gaat, is niet voldaan aan de eis van unidimensionaliteit. Het berekenen van een betrouwbaarheidscoëfficiënt voor een dergelijke toets is dan minder zinvol. Het is echter mogelijk dat bij een toets voor enkele onafhankelijke vaardigheden toch één faktor gevonden wordt omdat deze vaardigheden in gelijke mate afhankelijk zijn van de meer of mindere mate waarin de studenten kennis hebben genomen van de stof. Een dergelijke interpretatie van unidimensionaliteit is echter wel erg oneigenlijk. Overigens moet opgemerkt worden dat bij alle beoordelingen van studieresultaten waar men slechts één cijfer geeft, gehandeld wordt alsof er sprake is van unidimensionaliteit. Als alle items niet eenzelfde faktor zouden meten, zou het namelijk weinig zin hebben maar één cijfer te geven.

We vergelijken nu het toetsen van onderwijsdoelstellingen op het nivo van kennisreproductie met toetsen over doelstellingen op de hogere cognitieve nivo's. Op het nivo van kennisreproductie valt niet te verdedigen dat de verschillende kennisitems een samenhangende vaardigheid of een samenhangend geheel van vaardigheden vormen. De verschillende items staan op zichzelf. De kennisonderdelen worden ook afzonderlijk stuk voor stuk in het onderwijs aangeboden. Het betrouwbaarheidsmodel is niet van toepassing bij toetsen op dit nivo. Bij toetsen over doelstellingen op de hogere cognitieve nivo's is veel eerder sprake van een samenhangend geheel van vaardigheden; generalisatie en transfer tussen de verschillende delen ligt veel meer voor de hand. Een wat andere ingang op de betrouwbaarheidstheorie kan het verschil verder verduidelijken. Uitgangspunt bij de betrouwbaarheidstheorie is de opvatting dat een toets, c.q. de toetsitems, een steekproef vormen uit een populatie van alle mogelijke toetsen. De betrouwbaarheidscoëfficiënt zegt dan iets over de mate waarin de scores bij herhaalde meting met een vergelijkbare andere toets, gereproduceerd zouden worden. Bij toetsen bedoeld voor het meten van kennisreproductie gaat het om eindige populaties van items, die eigenlijk alle in de toets opgenomen zouden

* Bij zogenaamde 'onbenullige' items (vergelijk Hofstee, 1973) zal kwadrant IV leeg zijn.

moeten zijn. Vergelijkbare *andere* (parallele) toetsen over dezelfde stof zijn er dan niet. Bij toetsen op het nivo van de hogere cognitieve doelstellingen zijn de populaties niet eindig. Een regel (rule) bijvoorbeeld kan op talloze manieren toegepast worden. Enerzijds betekent dit dat hij met talloze verschillende voorbeelden uitgelegd kan worden, anderzijds zijn er vele mogelijkheden om verschillende items over dezelfde regel te formuleren. Konkluderend: in tegenstelling tot de toetsen op het nivo van kennisreproductie is er bij toetsen over hogere cognitieve doelstellingen weinig reden de geldigheid van het klassieke betrouwbaarheidsmodel te betwisten. Het is immers duidelijk wat onder parallele, vergelijkbare toetsen verstaan kan worden. Daarnaast kunnen de items door het optreden van transferprocessen beantwoord worden op grond van een samenhangend geheel van vaardigheden.

Een ander aspekt van de toepasbaarheid van de klassieke testtheorie op mastery toetsen is door Livingstone (1972) onderzocht. Volgens de klassieke testtheorie gaat het om het meten van verschillen in vorderingen tussen de studenten; het is immers de bedoeling onderwijsbeslissingen te nemen, zoals slagen of zakken, op grond van de verschillen tussen de studenten. Bij mastery toetsen wordt benadrukt dat het niet gaat om verschillen tussen studenten maar om afwijkingen van een criterium score. Livingstone heeft geprobeerd daarvoor een nieuwe testtheorie te ontwikkelen. Het resultaat van zijn werk was een theorie die zeer sterk op de klassieke theorie lijkt, maar een meer algemene vorm ervan is. Livingstone stelt namelijk dat in de klassieke testtheorie niet de variantie om de gemiddelde score genomen moet worden, maar de variantie om de criterium score. Op de klassieke manier is nu bijvoorbeeld de KR-20 betrouwbaarheidscoëfficiënt te berekenen, maar nu met varianties om de criterium score. Het resultaat is een nieuwe betrouwbaarheidscoëfficiënt die evenredig is met de klassiek berekende KR-20, maar ook met de afstand van de criterium score tot de gemiddelde score. De volgens Livingstone berekende be-

trouwbaarheidscoëfficiënt geeft dus een indicatie over de mate waarin de scores bij een herhaalde toetsing gereproduceerd kunnen worden, dit overeenkomstig de klassieke betrouwbaarheidscoëfficiënt. Deze informatie wordt bij Livingstone echter gekombineerd met informatie over de afstand tussen de scores en de criterium score. De volgens Livingstone berekende betrouwbaarheidscoëfficiënt zal dan ook toenemen naarmate de criterium score verder van de gemiddelde score ligt. Het lijkt duidelijker beide soorten informatie te scheiden. De informatie over de reproduceerbaarheid kan dan bepaald worden met de klassieke betrouwbaarheidscoëfficiënt, terwijl de informatie over de afstand van de scores van de criterium score in de frekwentieverdeling af te lezen is.

4.3. *De cesuur bij mastery toetsen.*

Als er één type onderwijs is waar absolute normen gehanteerd moeten worden dan is dat bij mastery onderwijs. Van tevoren is het mastery criterium waaraan iedere student moet voldoen, vastgelegd. Een relatieve norm lijkt niet op zijn plaats.

In het voorgaande is onderscheid gemaakt tussen doelstellingen op het nivo van kennisreproductie en de hogere cognitieve doelstellingen. Op het nivo van kennisreproductie levert het vaststellen van een absolute norm geen probleem. Alle items moeten goed beantwoord worden; eventueel wordt een enkele vergissing nog toegestaan.

Voor hogere cognitieve doelstellingen werd aangegeven dat transfer eerder regel dan uitzondering is. Items op dit nivo zullen moeilijk of minder moeilijk zijn, afhankelijk van de overeenkomst tussen het item en de behandelde stof. De P-waardes zullen in veel gevallen lager dan 100% zijn. Voor de mastery norm betekent dit dat nauwelijks geëist kan worden dat alle of vrijwel alle items goed beantwoord moeten zijn. Dit sluit echter het hanteren van een absolute norm nog niet uit. Een absolute norm, bijvoorbeeld 50% van de items goed geeft een voldoende, is

zeer goed mogelijk. Het bepalen waar die norm te leggen is echter bij items voor de hogere cognitieve doelstellingen moeilijk. Wanneer de items dicht bij de onderwezen stof blijven, lijkt het niet moeilijk te eisen dat deze items grotendeels goed beantwoord moeten worden. Wanneer de items nieuwe verderwegliggende problemen stellen, wordt het voor de docent steeds moeilijker te takseren of de transfer van de bestudeerde stof naar het te beantwoorden item niet een aantal ongekende problemen oproept. De empirie wijst uit dat docenten nauwelijks aan kunnen geven hoe moeilijk een bepaald item is (Mellenbergh, 1972). Hiermee is een, weliswaar negatieve, argumentatie voor het hanteren van een relatieve norm ook voor mastery toetsen gegeven. Positieve argumentaties voor relatieve normen zijn te vinden bij Wijnen (1972).

Literatuur

- Block, J. H., ed. *Mastery learning, theory and practice*. Holt, Rinehart and Winston, 1971.
- Bloom, B. S., J. T. Hasting, G. F. Madaus, *Handbook on formative and summative evaluation of student learning*, McGraw Hill, New York, 1971.
- Carroll, J. B., The concept of learning for mastery (in *Mastery learning*, Block, 1971).
- Groot, A. D. de, Ieder kan op den duur alles leren; gaat u ook nooit iets boven de pet?, RITP-memorandum 027, *Onderzoek van Onderwijs*, 9-10, 1973.

- Hofstee, W. K. B., Onbenullige toets items, *Nederlands Tijdschrift voor de Psychologie (DPO)*, 28, 189-198, 1973.
- Kramers-Pals, H., A. Pilot, J. F. M. J. van Hout, Mastery learning, methode om beter greep te krijgen op het leren. *Onderzoek van Onderwijs*, 2, 3-7, 1973.
- Livingstone, S. A., Criterion-referenced applications of classical test theory. *Journal of Educational Measurement*, 2, 13-26, 1972.
- Mellenbergh, G. J., *Studies in studietoetsen*, proefschrift, Psychologisch Laboratorium Universiteit van Amsterdam, 1971.
- Merrill, M. D., R. C. Boutwell, Instructional development, methodology and research. *Review of Research in Education*, 1, 95-131, 1973.
- Morrison, H. C., *The practice of teaching in the secondary school*, Chicago, 1926.
- Naerssen, R. F. van, Voor- en nadelen van toetsen met bekende items, *Nederlands Tijdschrift voor de Psychologie (DPO)*, 28, 259-266, 1973.
- Warries, E., Standaard beheersingscurves, *Memoereks onderwijsresearch*, 1, 1-4, 1973.
- Wijnen, W. H. F. W., *Onder of boven de maat*, Groningen, 1971.

Curriculum vitae

A. E. N. Rommes (geb. 1941) studeerde psychologie aan de Universiteit van Groningen; momenteel werkzaam bij het Centrum voor Didactiek en Onderzoek van Onderwijs van de T.H. Twente

Adres: CDO, T.H. Twente, Postbus 217, Enschede.