

De waarde van resumerende en tweede hands informatie

Inleiding

Informatie over empirisch onderwijskundig onderzoek is vaak resumerend of tweede hands, d.w.z. afkomstig uit niet algemeen beschikbare bronnen.

Ter ondersteuning van zijn theorie over de taxonomie van het cognitieve domein en mastery learning worden

door Bloom onderzoeken aangehaald, die slechts op aanvraag o.a. op microfilm in de V.S. te verkrijgen zijn. Prof. Freudenthal heeft een vijftal van deze (ongepubliceerde) studies opgevraagd en komt na bestudering tot de conclusie dat de betreffende onderzoeken de theorie van Bloom niet ondersteunen en bovendien door Bloom onjuist of onvolledig worden weergegeven. (*De redactie*)

Leerhiërarchische validatie van taxonomieën

1. Inleiding

Wat taxonomie van leerdoelen betekent, mag als bekend verondersteld worden. Baanbrekend was het werk, uitgegeven door B. S. Bloom¹ met zijn zes hoofdklassen 1.00 Knowledge, 2.00 Comprehension, 3.00 Application, 4.00 Analysis, 5.00 Synthesis, 6.00 Evaluation² die nog verder onderverdeeld werden, bijv. 1.10 Knowledge of specifics met de onderdelen 1.11 Knowledge of terminology, 1.12 Knowledge of specific facts enz.

Reeds in de oorspronkelijke publicatie werd gesteld of althans gehoopt dat de taxonomie een hiërarchische orde voorstelt, die leerhiërarchisch te valideren zou zijn (blz. 18):

Zoals we ze gedefinieerd hebben, maken de leerdoelen in een klasse zeer waarschijnlijk gebruik van en zijn gebaseerd op de gedragingen in de voorafgaande klassen van de lijst.

In de omvangrijke taxonomische literatuur wordt vrij algemeen het hiërarchisch karakter van een taxonomie verondersteld – althans als iets dat voor leerhiërarchische validering vatbaar is, soms ook als iets dat leerhiërarchisch is gevalideerd.

Leerhiërarchieën vallen vooral in de wiskunde, met haar deductieve structuur, op. Maar juist de wiskundige zal leerhiërarchische interpretaties van taxonomieën wantrouwen. Taxonomieën van de wiskunde nodigen ertoe uit ingevuld te worden volgens een hiërarchische structuur van de wiskunde, die vaak genoeg lijnrecht tegengesteld is aan een leerhiërarchie (zie hieromtrent³).

2. Bloom over Airasian

Volgens de literatuur is een leerhiërarchische validatie van taxonomie door Airasian geschied. B. S. Bloom zegt hieromtrent⁴ (blz. 23, 25):

In onderzoek waarover Airasian (1969) bericht, poneerde hij dat de elementen die weinig meer dan het leertype *Knowledge* vereisten, voor leerlingen gemakkelijker zouden zijn dan die *Comprehension* vereisten, en dat de moeilijkste leertypen die zouden zijn waarbij *Application* en *Analysis* te pas kwamen. Formatieve toetsen, die items op elk van deze niveaus bevatten, toegepast op verschillende klassen van leerlingen, steunden in 't algemeen deze hypothesen, d.w.z. de moeilijkheidsgraad van de vragen was, weinig uitzonderingen daargelaten, een steun voor de mening dat verschillende leerniveaus in 't spel waren, van *Knowledge* tot *Application* en *Analysis*.

Een tweede type van validatie van de elementen in een leertaak was eveneens door Airasian (1969) geleverd. Leerstof-deskundigen waren overeengekomen dat zekere elementen niet begrepen of geleerd konden worden, tenzij de leerling het met andere elementen in de taak had klaargespeeld, d.w.z. zekere elementen waren noodzakelijk voor het leren van meer complexe elementen. Wederom werden in verschillende klassen de geponeerde relaties tussen de elementen bevestigd door een analyse van de antwoord patronen in de toetsen die deze elementen bevatten. In 't geheel was het zo dat 75% van de antwoorden de geponeerde relaties staaften, terwijl 25% dit niet deden. Nader onderzoek is vereist om de voorwaarden vast te stellen waaronder de resultaten van

de a priori geponeerde afwijken. Desondanks grondvest dit soort validatie de wetmatigheid van de relaties tussen de elementen in een leertaak.

Het geciteerde verbaasde me uitermate. In de eerste alinea constateerde ik een gemis aan cijfers die er in de tweede alinea wel waren. Bovendien achtte ik – althans als wiskunde bij het onderzoek betrokken was – een dergelijke validatie uitgesloten om redenen die ik op 't eind van sectie 1 heb uiteengezet.

Wat de tweede alinea aangaat was ik omgekeerd verrast door de lage confirmatiegraad van 75%. Ik zou – ook weer als het om wiskunde gaat – vakdeskundigen toevertrouwen dat ze leerafhankelijkheden met een zekerheid van nagenoeg 100% voorspellen – uiteraard als er een deugdelijk toetsinstrument aan te pas komt.

Bloom beroept zich op Airasian's ongepubliceerde proefschrift (University of Chicago). Ik vraag me af of anderen, die zich op Airasian beroepen, deze bron ooit geraadpleegd hebben. Ik heb me door de Joseph Regenstein Library een microfilm van het proefschrift laten zenden⁵. Ik heb de microfilm inmiddels bestudeerd. Allereerst bleek dat Bloom's voorstelling van zaken, zoals boven vermeld, onjuist was.

Ten eerste, Airasian toetste niet Bloom's taxonomie. Comprehension en Analysis komen bij hem totaal niet voor in tegenstelling met wat Bloom ten deze stelt. In plaats hiervan introduceerde Airasian een niveau 'Process' dat, zoals het gedefinieerd en geoperationaliseerd werd, de bizarste gevolgen had, die ik ooit in verband met Taxonomie tegenkwam; tussen 'Process' en 'Application' had hij nog het niveau Translation.

Ten tweede heeft Airasian geen taxonomische hiërarchieën door middel van leerhiërarchieën gevalideerd, zoals Bloom in de eerste alinea suggereert. Hij deed *uitsluitend* hetgeen Bloom in de tweede alinea kenschetst als 'een tweede type van validatie': het bevestigen of weerleggen van hypothesen t.a.v. hiërarchieën in leerafhankelijkheden door middel van toetsen. Het lage percentage van overeenstemming (75%) behoeft geen nader onderzoek ter verklaring (zoals Bloom meent). Het heeft geen enkele objectieve betekenis, maar is de resultante van twee elkaar tegenwerkende fouten, zoals straks uiteengezet zal worden.

3. Airasian's instrument

Er werden twee leergebieden gekozen, 'highschool' scheikunde en 'first year algebra', en op deze gebieden twee opeenvolgende hoofdstukken in – niet nader gespecificeerde – 'veel gebruikte' schoolboeken. Deskundigen hadden de taak in deze leerstof de nieuwe leeritems te identificeren, hun taxonomische niveaus te bepalen en leerhiërarchische afhankelijkheden tussen de diverse items te poneren – ze waren voor deze taak gedurende zes weken geschoold. Bij deze leeritems werden toetsen geconstrueerd; de toetsconstructeurs waren er, naar het schijnt, niet van op de hoogte dat en welke leerhiërarchieën getoetst moesten worden. De toetsinstrumenten werden

daarna toegepast op overeenkomstige 'highschool' klassen.

Jammer genoeg bevat Airasian's proefschrift geen lijsten van de leeritems, noch van de taxonomische niveaus, noch van de geponeerde hiërarchische afhankelijkheden. Wel zijn de toetsinstrumenten aanwezig en uit deze alsmede uit de statistische gegevens valt het een of ander te gissen. Ik heb me hierbij en bij de nadere analyse tot het algebra-instrument beperkt.

4. Airasian's taxonomie

Airasian onderscheidt – afwijkend van Bloom – de zes gedragsniveaus:

1. Knowledge of terms, 2. Knowledge of facts, 3. Knowledge of rules and principles, 4. Process, 5. Translation, 6. Application. Ik licht deze terminologie toe aan de hand van het – fictief – voorbeeld vierkantswortel. Ik heb getracht dit zo getrouw als mogelijk te doen.
1. Definitie van vierkantswortel is Knowledge of terms.
2. Weten dat een positief getal een vierkantswortel bezit, is Knowledge of facts.
3. De regel $\sqrt{ab} = \sqrt{a} \sqrt{b}$ kennen is Knowledge of rules and principles.
4. De vierkantswortel $\sqrt{200} (= 10\sqrt{2})$ uit kunnen rekenen, is Process.
5. Een tekst uit omgangstaal in wiskundetaal – of omgekeerd – overbrengen (dus bijv. 'de vierkantswortel uit' in ' $\sqrt{\quad}$ ') is Translation.
6. De regel $\sqrt{ab} = \sqrt{a} \sqrt{b}$ kunnen toepassen is Application.

Zoiets is: leerhiërarchieën op zijn kop zetten. Er was zes weken training – of veeleer hersenspoeling – voor nodig om vakdeskundigen op dit averechtse spoor te zetten. Bij gebrek aan gegevens valt alleen uit het toetsinstrument na te gaan in welke mate dit gelukt is.

Verbale definities op 't laagste niveau plaatsen is een algemene taxonomistische gewoonte, die zich – zeker in de wiskunde met haar formele definities – averchts uitwerkt. Wiskundige leerprocessen starten nooit met definities, en formele definities kunnen helemaal aan het eind van leerprocessen en hoog in de leerhiërarchie staan. Algemene regels gaan in het wiskunde leren vaak niet aan hun toepassingen vooraf, maar volgen eruit.

Een toetsconstructeur die van de bedoelingen op de hoogte was geweest, had veel recht kunnen zetten. Maar telkens waar bij een leeritem Knowledge of terms stond, ging hij een formele definitie vragen volgens een van de twee schema's (met de vierkantswortel toegelicht):

het getal dat gekwadrateerd a oplevert heet de . . . van a ,

of
de vierkantswortel uit a is het getal dat . . .
waartussen zelfs nog een belangrijk niveau-verschil bestaat.

5. Airasian's methode van verwerking

De deskundigen hadden afhankelijkheden tussen leeritems zoals

A hangt af van B hangt af van C
geponeerd. Een leerling die volgens dit patroon scoorde (dus bijvoorbeeld A fout en B en C goed beantwoordde), werd geacht het patroon te bevestigen. Indien, zeg, 60% van de leerlingen een patroon bevestigden, werd er van 60% bevestiging van het patroon gesproken. Het door Bloom geciteerde percentage van 75% is het gemiddelde over alle in aanmerking genomen hiërarchieën.

Deze aanpak is foutief om twee redenen:

Ten eerste: 100% bevestiging van bovengenoemde 'afhankelijkheid' zou tot stand kunnen komen doordat alle leerlingen alle drie toetsitems A, B, C goed doen of alle drie fout doen, of doordat alle leerlingen B en C goed doen en geen A goed doet (en wel niet omdat er enige afhankelijkheid tussen de items zou bestaan, maar omdat B en C te gemakkelijk zijn en A te moeilijk is). Alle deze en hierop gelijkende gevallen mogen voor het bevestigen van afhankelijkheden niet meetellen, als zijnde statistisch insignificant. Er is een hele theorie voor nodig om de significantie van patroonbevestigingen te bepalen met als essentieel element *voorwaardelijke* moeilijkheidsgraden (het goed beantwoorden van A onder de voorwaarde van goed beantwoorden van B en C, enz.).

Ten tweede: 60% bevestiging van bovengenoemde afhankelijkheid zou tot stand kunnen komen doordat, zeg, A en C door 100% en B door 60% van alle leerlingen goed wordt gedaan. Het is dan onjuist om te stellen dat

A hangt af van B hangt af van C
voor 60% bevestigd wordt. Indien enige conclusie mogelijk is, dan is het hetzij dat

B hangt af van A hangt af van C
hetzij dat

B hangt af van C hangt af van A
voor 100% wordt bevestigd.

Deze fout had kunnen worden voorkomen door niet losse hypothesen maar hypothesen tegen elkaar te toetsen.

6. Airasian's resultaten

Voorzover de schaarse gegevens, door Airasian verstrekt, conclusies toelaten, zijn de afzonderlijke uitspraken omtrent de bevestiging van geponeerde leerhiërarchieën ge-

heel of nagenoeg geheel bepaald door de twee in sectie 5esignaleerde fouten.

Het taxonomistische op de kop zetten van de leerhiërarchieën, waarin de deskundigen waren gecoacht, had in principe kunnen leiden tot een bevestigingsgraad van 100% voor concurrerende i.p.v. de geponeerde hiërarchieën. Dankzij de fout ad 2 werden in dergelijke gevallen nog bevestigingsgraden van om en bij de 50% geconstateerd. Dankzij de fouten ad 1 werden in andere hiërarchieën insignificante bevestigingsgraden van bij de 100% behaald en ten onrechte meegeteld. Hierdoor werd het bevestigingspercentage flink opgestoten, tot de door Bloom vermelde 75% toe.

7. Conclusie

Het onderzoek van Airasian betekent niets voor de leerhiërarchische validatie van taxonomie. Het is in zichzelf van nul en gener waarde. Om dit in te zien is geen diepgaande analyse vereist. Hoewel de – theoretische en statistische – fouten apert zijn, zal ik er in een uitvoerige publicatie elders⁶ nog gedetailleerd op ingaan. Vermoedelijk heeft niemand eerder deze ongepubliceerde dissertatie bekeken.

Noten

1. B. S. Bloom, ed., *Taxonomy of Educational Objectives. I. Cognitive Domain*, 1956.
2. Ik geef er de voorkeur aan deze uitdrukkingen niet te vertalen.
3. H. Freudenthal, *Structuur der wiskunde en wiskundige structuren*, *Ped. Stud.* 1979 (56), 51-60.
4. B. S. Bloom, *Human Characteristics and School Learning*, New York, McGrawHill, 1976, XII + 284 p.
5. P. W. Airasian, *Formative evaluation instruments: A construction and validation of tests to evaluate learning over short time periods*, 1969. VIII + 193 p. Nummer microfilm T 17387. Het ontvangen exemplaar berust op het IOWO, Tiberdreef 4, Utrecht.
6. Ways to report on empirical research in education. Verschijnt in: *Educational Studies on Mathematics* 10, afl. 4.

Mastery learning

1. In een kortgeleden verschenen boekje¹ breekt B. S. Bloom een lans voor Mastery Learning. Hij verstrekt indrukwekkende grafieken die de geweldige superioriteit van Mastery Learning boven 'more conventional approaches' moeten demonstreren. De grafieken hebben betrekking op vijf experimentele onderzoeken: Second Language (Binor), Imaginary Science (Arlin), Matrix Algebra (Block), Matrix Algebra (Anderson), Elementary Probability (Le-

vin).

Het valt in Bloom's relaas op dat wel een groot aantal verfijnd verwerkte statistische daten worden vermeld, maar dat de meest elementaire gegevens ontbreken. Er is niets over steekproefgrootte, vergelijkbaarheid van experimentele en control-groep, leerjaar, leerstof (normaal of experimenteel), niveau van de leerstof (routine of hoger), duur van de experimenten (normale schoolperiodes of

experimentele), de aard van het control-onderwijs, retentie enz. Vooral valt bij een van die grafieken (Arlin) het ontbreken van het 'summative achievement' op.

Alle vijf onderzoeken worden door Bloom als ongepubliceerde dissertaties van University of Chicago geciteerd. Ik heb me door de Joseph Regenstein Library microfilms van vier van deze proefschriften laten zenden; Binor's proefschrift kon aldaar niet worden achterhaald.²

Als volgt een kort verslag van mijn bevindingen, waarover ik elders³ in extenso zal berichten.

2. Geen van de door Bloom afgedrukte grafieken komt in de geciteerde bronnen voor. In twee gevallen (Block, Arlin) laat Bloom's grafiek zich vermoedelijk uit gegevens in de bronnen reconstrueren, in een van deze gevallen (Arlin) echter pas na correcties, ten dele ten laste van Arlin, ten dele van Bloom. In een derde geval (Anderson) zijn er geen gegevens in de bron, die Bloom's grafiek rechtvaardigen, in het vierde geval (Levin) zijn de gegevens in de bron in strijd met Bloom.

3. Het gegeven 'Summative achievement' in de grafiek Arlin ontbreekt al in de bron. Statistische gegevens in de bron doen het vermoeden rijzen dat de Mastery en de control-groep in het 'summative achievement' niet significant verschilden (zie Appendix B; rubriek Chap. 8, original % correct).

4. In alle experimenten werd met zeer kleine steekproeven gewerkt (10-25). De experimenten waren van zeer korte duur (4-12 lessen). De leerstof was ad hoc geconstrueerd, d.w.z. niet normaal. De vlag 'Matrix Algebra' dekte de lading totaal niet; het was puur rekenen of zeer elementaire algebra en voor de betrokken leerlingen geen nieuwe leerstof. De proeve van leerstof van 'Imaginary Science' is geheel onbegrijpelijk; het lijkt geen leerstof, maar veeleer iets in de geest van 'Monopoly'. 'Elementary probability' is serieuze leerstof.

5. Bij geen van de onderzoeken staat het vergelijken van Mastery en non-Mastery voorop. Er is derhalve weinig aandacht geschonken aan een redelijke definitie van het control-onderwijs. Het was eenvoudig: uit het ad hoc geconstrueerde Mastery onderwijs Mastery elementen

schrappen, bijv. formatieve tussentoetsen wel toedienen, maar geen correcties aan de leerlingen doen toekomen. Zonder nader onderzoek bestaat er geen aanleiding dit soort control onderwijs aan normaal onderwijs gelijk te stellen. Op zijn minst twee van de auteurs (Block, Levin) zijn dan ook erg voorzichtig in het trekken van conclusies wat de superioriteit van Mastery Learning aangaat.

6. In het meest waardevolle onder deze onderzoeken (Levin) bleek de experimentele groep het alleen op de routine-taken significant beter te doen dan de control-groep. Dit feit wordt door Bloom niet vermeld.

7. Naar uit het voorafgaande blijkt, is Bloom's rapportage vrij ongenueanceerd. Gegevens, die relevant zouden zijn voor het herhalen van dergelijke experimenten, worden niet verstrekt. Dit treft te meer daar het hier gaat om onderzoek dat niet gepubliceerd en dus moeilijk toegankelijk is.

H. Freudenthal

Noten

1. B. S. Bloom, *Human Characteristics and School Learning*, New York, McGrawHill 1976, XII + 284 p. Speciaal blz. 55-56.
2. Anderson, L. W., *Time and school learning*. 1973. VI + 128 p. T 24567. Arlin jr., M. N., *Learning rate and learning rate variance under mastery learning conditions*. 1973. VII + 206 p. T 24704. Block, J. H., *The effects of various levels of performance on selected cognitive, affective, and time variables*. 1970. VII + 331 p. T 22282. Levin, Tamar., *The effect of content prerequisite and process-oriented experiences on application ability in the learning of probability*. 1975. X + 216 p. T 25619. De vermelde nummers zijn de microfilm-nummers. Copieën van de microfilms berusten op het IOWO, Tiberdreef 4, Utrecht.
3. *Ways to report on empirical research in education*. Verschijnt in: *Educational Studies on Mathematics*, 10, afl. 4.