

Kort weerwoord op K.M. Stokking: Evaluatie en implementatie

J. H. SLAVENBURG

Stokking (1987) onderscheidt in zijn bespreking van de eerste evaluatieresultaten van het OSM-project een viertal aspecten: het vooraf aangeven van exacte evaluatiecriteria, het trekken van conclusies met behulp van causale modellen, de aandacht voor de implementatiewaarde van de programma's en de negatieve eindconclusies. Ik geef graag een kort weerwoord op het door Stokking bij elk van deze aspecten geformuleerde commentaar, met name op die punten waarop ik met hem van mening verschil.

1 *Exacte evaluatiecriteria*

Stokking stelt dat in het OSM-project weliswaar exacte evaluatiecriteria geformuleerd zijn, maar dat op tal van punten toch achteraf interpretatie noodzakelijk bleek en soms water bij de wijn werd gedaan.

Ik zou graag willen volhouden dat in elk geval met betrekking tot de exacte evaluatiecriteria geen water bij de wijn gedaan is: de (tenminste) landelijk gemiddelde prestaties op genormeerde schoolvorderingstoetsen zijn als criteria gekozen en gehanteerd.

2 *Causale modellen*

Stokking stelt dat geavanceerde technieken als Lisrel kennelijk gewoon niet zo geschikt zijn voor onderzoek in levensechte situaties. Hij stelt verder dat men zich kan afvragen of het OSM-verklaringsmodel wel geschikt is voor het opsporen van de oorzaken van het gebrek aan effect van de OSM-programma's.

Ik ben het met Stokking eens dat lineaire

modellen met complexe schattings- en toetsingsprocedures aanleiding geven tot conceptuele problemen (in het bijzonder met betrekking tot de betekenis van allerlei relaties tussen de modelvariabelen) en tot realiteitsproblemen (modellen die goed passen op de verzamelde gegevens zijn zeldzaam). Echter, deze modellen bieden ook de mogelijkheid, juist in levensechte situaties, allerlei storende variabelen enigszins onder controle te krijgen (door ze te meten en hun invloed in het model te schatten), een controlemogelijkheid die veelal niet of nauwelijks door de gebruikelijke quasi-experimentele proefopzetten wordt geboden (met uitzondering van het regressie-discontinuïteits design). Het zoeken naar realistische schattings- en toetsingsprocedures voor deze modellen is nog in volle gang (cf. Lohnes, 1986). Zij kunnen zeker nog niet als onbruikbaar worden afgeschreven. Anderzijds: de (geringe) mate waarin de modellen bij de waarnemingen passen zegt ook iets van de mate waarin wij in staat zijn de werkelijkheid in onze theorievorming te beschrijven (Stokking noemt zelf al een voorbeeld: aanwezigheidstijd van de leerlingen) en ook – wellicht nog meer – van de mate waarin wij erin slagen de modelvariabelen (vooral) valide en (ook) betrouwbaar te meten.

Het aangeven waardoor de OSM-programma's niet meer effect hadden dan zij hadden is geen onderwerp van onderzoek geweest in mijn proefschrift. Overigens: dit type van vragen is per definitie niet door middel van onderzoek te beantwoorden. Het antwoord kan – en dan nog alleen met gebruikmaking van een causaal model – eventueel gesimuleerd worden. Ik heb gesteld dat het mij – gezien het toch bepaald zwakke verband tussen de programma-implementatiewaarden en de prestaties op de genormeerde schoolvorderingstoetsen – niet zinvol voorkwam een dergelijke simulatie uit te voeren. (In het binnenkort verschijnende

proefschrift van Van Tilborg zijn dergelijke simulaties overigens wèl opgenomen.)

De vraag die ik heb proberen te beantwoorden luidt ook anders dan Stokking zegt: door welke variabelen (=oorzaken in een passend causaal model!) worden de (lager dan verwachte) toetsprestaties verklaard? Het antwoord op deze vraag leidt tot het rijtje variabelen dat door Stokking is opgesomd.

3 Implementatie

De verklaringswaarde van programma-implementatie wordt, in tegenstelling tot hetgeen Stokking stelt, ook op pagina 55 als belangrijk gekenschetst. Alleen wordt daàr gesteld dat dit niet de reden is waarom dit onderwerp op pagina 55 aan de orde is (de verklaringswaarde van deze en andere variabelen wordt namelijk in hoofdstuk 4 besproken). Op pagina 55 is de implementatiewaarde aan de orde als beschrijving van de mate waarin het te evalueren programma wordt uitgevoerd zoals bedoeld. Als een programma heel anders dan zoals bedoeld wordt uitgevoerd zijn er geen generaliserende conclusies over dat programma zoals bedoeld te trekken.

Het weergeven van de gegevens over de doeleffectiviteit van de programma's gebeurt in eerste instantie voor de volledigheid en in tweede instantie om programma's die qua effect 'iets voorstellen' te selecteren voor opname, via hun implementatiewaarde, in de causale modellen. De gegevens over doeleffectiviteit worden dus niet gepresenteerd vanwege hun belang voor de implementatiewaarde.

Terzijde – gezien de betreffende opmerkingen van Stokking – ik heb een begeleidingsmodel bruikbaar genoemd als het zelf voldoende geïmplementeerd was, als het ermee begeleide programma voldoende geïmplementeerd was en als tussen beide implementatiewaarden een positief causaal verband bestond.

4 Eindconclusie

Als het doel van de OSM-programma's is het realiseren van tenminste de landelijk gemiddelde prestaties op bepaalde schoolvorderin-

gentoetsen en als geconstateerd moet worden dat dit doel niet gerealiseerd is, kan ik niets bedenken waarom de uitkomsten op de drie gebruikte leestoetsen niet de door Stokking gewraakte doorslag zouden mogen geven. Hij voert in dit verband zelf ook geen argumenten aan.

Het achterblijven bij het landelijk gemiddelde had, naar Stokking stelt, gerelateerd moeten worden aan de achterstand die voortvloeiend uit de inbreng van de OSM-programma's reeds aanwezig was. Naar mijn mening is de in deze stellingname besloten onderzoeksvraag beantwoord door te kijken naar de prestaties van de quasi-controlegroep: die waren meestal beter dan die van de OSM-groep.

Stokking grijpt zich ten aanzien van de effecten van de OSM-programma's vast aan enkele (dunne) strohalmen: de (geringe) positieve causale relatie tussen de implementatiewaarden van de OSM-programma's en de toetsprestaties en die tussen de implementatiewaarden van de begeleidingsmodellen en de onderwijsstimuleringsprogramma's. Het verbeteren van de implementatie van de begeleidingsmodellen zal leiden tot verbetering van de implementatie van de onderwijsprogramma's en dit zal leiden tot verhoging van de toetsprestaties. Een plausibele bewering waarmee ik kan instemmen. Echter, een even plausibele bewering die waarschijnlijk – gezien de geconstateerde sterke verbanden – tot veel grotere effecten kan leiden is die met betrekking tot de beïnvloeding van de intelligentie, de werkhouding e.d. De vraag is alleen: hoe. Om die vraag te kunnen beantwoorden is nog veel onderzoek en ontwikkelingswerk noodzakelijk. Tot dat gebeurd is, zullen wij van de door Stokking bedoelde strohalmen gebruik moeten maken.

Literatuur

- Lohnes, P.R., Correlated factors modeling via weak mathematics. *American Educational Research Journal*, 1986, 23, 289-302.
- Stokking, K.M., Evaluatie en implementatie. *Pedagogische Studiën*, 1987, 64, 35-40.