

[Ik heb pas laat kennisgemaakt met het tentamenmodel van Van Naerssen. Ik zag dat aanvankelijk als een model voor hoe studeten zijn te manipuleren, en had veel tijd nodig om over die hobbel heen te komen. Maar het is gelukt. Met SPA-model is een verdere uitwerking van het idee van Van Naerssen voor een tentamenmodel. Het zou mijn proefschriftonderwerp worden, maardat is weer een ander verhaal. bw 2025]

Bespreking van Van Naerssen A mathematical model for the optimal use of criterion referenced tests. [pdf van dat artikel: <https://benwilbrink.nl/publicaties/74vnaerssenmodel.pdf>]

Uitgangspunt is een volkomen transparante tentamensituatie (wat impliceert dat er van relatieve beoordeling op dat tentamen geen sprake kan zijn, zoals van vN ook opmerkt.

Natuurlijk kunnen de voor dit tentamen gestelde normen wel relatief zijn, gebaseerd op vroegere ervaringen e.d.)

Het gebruik van 'criterionreferenced' als term door van N. schijnt dan ook niet helemaal te dekken wat daar over het algemeen in de literatuur ongeveer mee bedoeld wordt; Hofstee zou zeggen dat ook in die literatuur de criterions uiteindelijk relatief bepaald zijn.

Belangrijke veronderstelling is dat bij zakken de student het tentamen over mag doen (ja zelfs, het tentamen over moet doen), en dat wel net zo vaak tot hij er voor slaagt. De bedoeling schijnt te zijn dat we het onderwijs zo inrichten dat studenten zoveel mogelijk de eerste keer al kunnen slagen (zie eerdere artikelen van van Naerssen). Anderzijds is het stellen van de eis van een behoorlijk hoog prestatie niveau op een of andere oncontroleerbare wijze veelal beschouwd als impliciete eis bij een onderwijs gebaseerd op de criterion referenced testing filosofie. En juist wat dit laatste punt betreft heb ik bedenkingen die min of meer ernstig kunnen zijn, niet tegen het model van Van Naerssen, maar tegen de wenselijkheid van hanteren van dat model in de onderwijssituatie. Het laten overdoen van onvoldoende gemaakte studieonderdelen moet in het algemeen toch als een verspilling van tijd en energie van de betrokken beschouwd worden. Verspilling, omdat moeilijk aantoonbaar is dat de tijdbesteding van het overdoen beter is dan alternatieve mogelijkheden om studie en doceertijd te investeren (we zien nog maar even af van de lange tijd die meestal gemoeid is met het afleggen van het tentamen zelf). Bij het laten overdoen van studieonderdelen wordt, zoals ik elders heb laten zien, de veronderstelling gemaakt dat er van een Aptitude Treatment Interactie sprake zou zijn. Verschillende studenten krijgen een verschillende behandeling, wat alleen zinvol is wanneer die handelwijze een hoger pay off geeft dan het geven van een gelijke behandeling aan alle studenten. Deze argumentatie is op te vatten als een tegenargumentatie voor het model van criterion referenced testing zoals dat op meer rigide wijze nogal eens verdedigd wordt door learning for mastery adepten. Natuurlijk is het vaak mogelijk herkansingen zo in te richten dat vrijwel geen tijd en energie voor de student verloren gaan, zo snel volgend op de eerste tentamenafname dat er van vergeten nog nauwelijks sprake kan zijn. En met dat laatste zitten we dan in een situatie waarvoor Van Naerssen zijn model helemaal niet heeft opgesteld. Het ziet er dus naar uit dat bij een uitwerking van deze argumenten een tegenstelling geconstrueerd kan worden tussen het model van Van Naerssen en de veronderstelling dat het laten overdoen van tentamens onder de omstandigheden waarin Van Naerssen zijn model wil toepassen verspilling van tijd en energie voor de student betekent. In het volgende zullen we deze tegenstelling niet verder uitwerken, omdat het allereerst er om te doen is de aard van het model van Van Naerssen te onderzoeken, en niet om meteen al tegenmodellen te ontwikkelen.

2. Learning curves and forgetting curves.

De aanpak van Van N. zoals in eerdere publicaties, is n.a.v. die eerdere artikelen door mij becommentarieerd. Voor van N. was het probleem met deze leercurven dat het fenomeen 'overlearning' er niet in te vangen was. Dit probleem wil hij nu ondervangen niet een model 'adopted from score theory. The true score is a (normal) ogive function of the 'underlying ability'. This 'test characteristic curve' is the sum of the 'item characteristic curves', that describe the probability of answering the item correctly as a function of the ability.

During the study phase the student moves on the ability dimension from left to right. This dimension is just what the examination measures. If the student ceases to study he moves to the left on the ability dimension."

"Because exact empirical material is lacking, the simplest assumptions are made, namely that the student is moving with a uniform velocity, to the right or to the left. This means that learning curve and forgetting curve both are supposed to be ogives, though the former is (much) steeper."

We zitten hiermee dan volop in de problemen.

Allereerst is het nog maar de vraag of het soort van leren en leerstof dat van N. hierbij voor ogen staat wel zo relevant is voor de (hoger) onderwijs situatie. De eerste gedachte die zijn veronderstelling oproept is dat het waarschijnlijk mogelijk is om een ruwe indeling in twee soorten leerstof te maken, die ieder een geheel andere 'acquisitie' en vergeet curve laten zien. Het leren van eenvoudige kennis zal kunnen verlopen volgens het alles of niets leren model, en deze kennis zal zeer snel weer vergeten worden en moeilijk weer opnieuw te leren zijn. Begripsmatige leerstof is moeilijk te verwerken, maar kan de eigenschap hebben bij een behoorlijk opgezet onderwijs dat de vergeetcurve negatief is, d.w.z. dat bij onverwachte toetsing pakweg een jaar later de score op de toets hoger uitvalt dan bij de eerste en eigenlijke tentamenafname. (Cook, in Lindquist 1951).

[Het bovenstaande verschil is op voor de hand liggende wijze te schetsen, dat heb ik maar niet gescand]

Op deze globale noties wat doorfilosofierend zouden we bijvoorbeeld kunnen bedenken dat begripsmatige leerstof tenminste tot een bepaald niveau beheerst moet worden, wil er van een 'groei' sprake kunnen zijn. in de figuur [niet gescand] zijn twee gevallen geschetst: in het ene geval is de stof onvoldoende geïntegreerd en wordt in de maanden volgend op het tentamen weer vergeten, in het andere geval is de stof voldoende geïntegreerd om er in het volgende onderwijs mee te kunnen werken en verder te integreren.

Cook heeft enkele onderzoeken genoemd waaruit dergelijke samenhangen blijken voor kennis versus begripsmatige leerstof. De vraag hier is iets anders: eigenlijk is de stelling dat begripsmatige leerstof niet begripsmatig verwerkt kan worden door de student, in welk geval een vergeetcurve gevonden zal worden zoals bij leerstof bestaande uit kennisitems gevonden wordt. Ik ben benieuwd of ook recenter onderzoek aanwijzingen in de richting van deze veronderstelde samenhangen zou kunnen geven.

Overigens noeft ook in het geschetste geval de implicatie voor studenten die het 'kritische' niveau van beheersing niet gehaald hebben, nog niet te zijn dat zij dit studieonderdeel over moeten doen totdat zij dat niveau wel halen. Ook dan maken we weer de veronderstelling dat er een disordinaire aptitude treatment interactie bestaat die

deze handelwijze rechtvaardigt.

Een dergelijk ATI hoeft echter geenszins te bestaan. Dat kan bijv. afhangen van de reden waarom de student het kritische niveau niet gehaald heeft, het kan ook afhangen van de inwisselbaarheid van verschillende studieonderdelen tegen elkaar, waarmee ik bedoel dat bij een onvoldoende gemaakt studieonderdeel A van de student gevraagd wordt om een alternatief studieonderdeel B of C te doen, dat niet gedaan hoeft te worden door degenen die een voldoende gehaald hebben, waarbij bovendien vakken B of C ook geheel andere inhoud kunnen hebben dan vak A. (Als B of C dezelfde inhoud als A zouden hebben, maar bijv. op een eenvoudiger niveau gepresenteerd, dan zitten we weer in dezelfde situatie als bij domweg laten overdoen wat de veronderstelling van ATI betreft).

Een eigenaardigheidje van de in de eerste figuur getekende leercurve is dat hij positief versneld is : het niveau van beheersing van begripsmatige stof gaat het snelst omhoog de laatste dagen voor het tentamen, wanneer ook het beste overzicht over de stof door de student verkregen wordt. Deze idee is zeer zwak gegrondvest in mijn persoonlijke ervaring, ik kan me geen onderzoek naar dit verschijnsel voor de geest halen, het zou echter wel degelijk een reëel fenomeen kunnen zijn al was het slechts in die zin dat het verschijnsel introspectief door een groot aantal studenten bevestigd zou worden (waarmee het belang er van is aangetoond, doch niet de realiteitswaarde).

Laten we nu overgaan tot de vaardigheidsdimensie, "This dimension is just what the examination measures." Ik ben bang dat van N. zich hier vreselijk in de nesten aan het werken is. Ik heb me altijd al afgevraagd wat tentamens nu eigenlijk meten. Van N. heeft zich daar ook wel eens over uitgelaten (bijv. onderwijscongres 1968). Dan blijkt toch altijd maar weer dat wat tentamens meten niet zoiets als beheersing van de stof is, maar een complex van ijver, tijdbesteding, intelligente vaardigheden, kwaliteit van het onderwijs e.d., een zootje dat bij factoranalyse van tentamenresultaten dan ook nogal eens slechts één grote factor op schijnt te leveren. Als dit dan allemaal zo zou zijn, en ik zou daar graag wat theoretische en empirische gegevens over willen hebben, wat is dan de zin van een tentamenmodel zoals van N. dat hier aan het formuleren is? Ik zie het niet, maar misschien komt dat nog. ik heb dus nogal wat moeite met de notie als zou de student op die vaardigheidsdimensie van links naar rechts of omgekeerd bewegen. Het probleem is natuurlijk op te lossen door de terminologie in te wisselen voor een andere met minder lastige connotaties.

Je zou er ook anders tegenaan kunnen gaan, vanuit de vraag bijv. wat we in dit verband onder meten moeten verstaan. Koppel dat aan het te verwachten verschijnsel dat onder de door van N. gepostuleerde tentamenomstandigheden (grote transparantie) de spreiding in tentamen scores klein zal zijn, en we hebben weer een aardig probleem, bij de hand.

De student beweegt op de vaardigheidsdimensie van de ene naar de andere kant (zonder noodzakelijkerwijs de uiteinden te bereiken). Ik heb moeite met de overgang van een latent ability dimension die over studenten gedefinieerd is (dat is toch zo?) naar een andere dimensie die geldt voor één bepaalde student. Ik zou dit soort stappen graag uitgewerkt zien.

Merkwaardigerwijs veronderstelt van N. tevens dat ook tijdens het leren weer vergeten plaats vindt, in tegenspraak met het alles of niets leren model dat hij voor het leerproces postuleerde waarin ieder eenmaal geleerd item ook onthouden blijft. Je vraagt je af wat dan de zin is van het postuleren van het alles of niets leren model, vooral ook omdat het in deze onderwijs situaties bar weinig uitmaakt of het alles of niets model of traditionele

modellen gehanteerd worden omdat ze beide dezelfde leercurve goed beschrijven.

3. The formula for learning and forgetting.

Van N. voert hier de volgende zes grootheden in:

G the engagement, the measure in which the student concentrates on the subject matter.

C the capacity to learn the specific subject matter

L the length of the subject matter

D the difficulty of this subject matter

I the isolatedness, the measure in which the subject matter is incoherent and not integrated in a larger part of knowledge

M the memory, the measure in which the student is able to remember what he has learned.

Velocity of forgetting is I/M

A^* is de verandering in vaardigheid

W is tijd

V is snelheid van leren

$$V = A^* \cdot W = GC/LD - I/M$$

Ik vind dit maar een raar stelletje grootheden, maar zal even afwachten welke grappen van N. er mee uit gaat halen

4. The units of time and ability.

"As is usual in test theory, it is assumed that true score is such an ogive shaped function of the ability that, as ability increases from minus infinity to plus infinity, true score increases from a basic value T_0 until the maximum value 1.

Van Naerssen refereert hier kennelijk aan de test characteristic curve, die gedefinieerd is over een populatie studenten. Dit staat dan weer in tegenstelling tot de verzameling van metingen op verschillende tijdstippen naar van deze de student, waaraan hij hier vervolgens refereert:

"Thus it may be better to estimate T_0 empirically, for instance by administering the test to a group which has not yet begun to study the subject matter."

Er is met de laatste uitspraak nog een andere moeilijkheid, dat Van N. niet consequent is door ook studenten die net aan de cursus beginnen niet in de gelegenheid te stellen de kennis en inzichten waarover ze mogelijk (latent) beschikken weer even op te frissen vóór de toets afname. Immers, dat is de procedure die hij voorstelt bij onverwachte toetsing enige tijd ná het eigenlijke tentamen, waar hij de studenten ook de gelegenheid meent te moeten geven om de stof nog even op te frissen omdat anders de meting niet geheel zinvol lijkt. Hoe dit ook precies ligt, in ieder geval is het van niet te verwaarlozen belang om rekening te houden met deze door

Van N. aangebrachte asymmetrie omdat er vertekende beelden uit zouden kunnen ontstaan bij empirisch onderzoek.

5. Latent ability.