

# Functionele betekenis van itembanking voor didactische doeleinden\*

G. TISTAERT

Katholieke Universiteit, Leuven

## Samenvatting

Reeds gedurende een tiental jaren bestudeert men in het Centrum voor Psychopedagogisch & Didactisch Onderzoek (C.P.D.O.) van de Leuvense Universiteit problemen in verband met didactische evaluatie. Itembanking omvat de studie van de instrumentalisering van operabele doelstellingen en het functioneel voorradig hebben van een 'groot' aantal vooraf bestudeerde items waarmee didactische doeleinden kunnen nagestreefd worden. Het hoofddoel dat men nastreeft met een itembank is het ter beschikking hebben van een ruim aantal doelstelling-valide items waarmee evaluatie-proeven, i.c. objectieve proeven voor didactische rendementsevaluatie, kunnen samengesteld worden. Naast deze hoofdfunctie kan men verder ook een instruerende en een coördinerende functie aanduiden.

## 1. Inleiding

Sedert ongeveer tien jaar worden in het Centrum voor Psychopedagogisch en Didactisch Onderzoek (C.P.D.O.) van de Universiteit te Leuven problemen bestudeerd die verband houden met didactische evaluatie. De concrete uitwerking, met inbegrip van empirisch onderzoek, concentreerde zich tot heden hoofdzakelijk op de produkt-evaluatie of de didactische rendements-evaluatie.

Door didactische rendementsevaluatie duidt men het gehele proces aan van systematisch en

methodologisch verantwoord verkrijgen, verwerken en interpreteren van informatie over de onderwijsleerresultaten op school via het resultaat dat de leerlingen bereiken die het onderwijs volgden. Om de gewenste informatie te verwerven heeft men instrumenten nodig. De ontwikkeling van deze instrumenten verloopt in verschillende fasen, die schematisch als volgt kunnen omschreven worden.

- a. De verwachtingen die men heeft t.o.v. het onderwijs worden uitgedrukt in operabele doelstellingen<sup>1</sup>.
- b. De leerdoelstellingen worden geïnstrumentaliseerd in doelstelling-valide items. D.w.z. men redigeert en bestudeert vragen, opgaven en taken waarmee men in de realiteit kan onderzoeken of de leerlingen de gewenste gedragsveranderingen, die in de leerdoelstellingen werden uitgedrukt, verworven hebben.
- c. Voor zover het aantal doelstellingen het gewenste aantal items van een proef overschrijdt, neemt men op aselechte wijze een al of niet gestratificeerde steekproef uit het universum van de doelstellingen en men stelt een evaluatieproef samen.
- d. Aan de hand van vooraf gestelde criteria interpreteert men de verkregen informatie in het licht van de gestelde doelen.

Itembanking situeert zich in de tweede fase van het evaluatieonderzoek en volgt chronologisch op het doelstellingsvraagstuk. De essentiële opdracht van itembanking bestaat in de studie van de instrumentalisering van leerdoelstellingen en het functioneel voorradig hebben van een groot aantal items waarmee didactische doeleinden kunnen nagestreefd worden.

\* Dit artikel verscheen eerder in het *Tijdschr. voor Opvoedk.* 1973-74 (19) nr. 3, 211-228.

Het zou natuurlijk mogelijk zijn, telkens als een evaluatieproef nodig is, een steekproef te nemen uit de inventaris van de geformuleerde leerdoelstellingen en deze te itemiseren. Wanneer men evenwel aan een periodieke en algemene evaluatie denkt, waarbij een dienstverlenend Centrum proeven ter beschikking kan stellen van scholen of groepen van scholen, die het rendement van het verstrekte onderwijs willen evalueren, dan is deze werkwijze noch economisch noch efficiënt. Daarom suggereerden sommige personen en instanties een voorraad van items aan te leggen over alle in de inventaris opgenomen leerdoelstellingen<sup>2</sup>.

Voor dergelijke verzamelingen van items worden in de literatuur twee verschillende termen gebruikt: itempool en itembank<sup>3</sup>. Wij duiden door de term itempool een verzameling van items aan die (nog) niet systematisch bestudeerd werden en waarvan dus geen parameters bekend zijn. De term itembank, c.q. didactische itembank behouden we voor een gestructureerde voorraad van items, waarvan de inhoudelijke en technische eigenschappen bestudeerd zijn en waarmee specifiek didactische doeleinden nagestreefd kunnen worden<sup>4</sup>.

In onderhavige bijdrage zullen wij in hoofdzaak de functionele betekenis van een didactische itembank aanduiden. Wij maken hierbij onderscheid tussen wat o.i. het hoofddoel is van een didactische itembank en wat men kan aanduiden als bijkomende functies. Tenslotte vermelden wij exemplarisch enkele voorlopers van het voorhanden hebben van goede items en noemen enkele actuele projecten in verband met itembanking<sup>5</sup>.

## 2. Doel en functie van een didactische itembank

Itembanking gebeurt met het oog op bepaalde doeleinden. Het hoofddoel van een itembank is het voorhanden hebben van een groot aantal goede items, waarmee proeven voor didactische rendementsevaluatie kunnen samengesteld worden. Naast deze hoofdfunctie kan een itembank ook voor bijkomende dienstverlening zorgen bij instructie en coördinatie.

## 2.1 Hoofddoel: samenstellen van evaluatieproeven

### 2.1.1. Omschrijving van objectieve proeven voor didactische rendementsevaluatie

Evaluatie in brede betekenis kan men omschrijven als een systematisch proces dat uit drie verschillende fasen bestaat: het verkrijgen, verwerken en interpreteren van informatie over het verloop van het didactisch handelen op school. Bij produkt-evaluatie hebben deze drie fasen betrekking op de mate waarin de leerlingen de leerdoelstellingen verworven hebben. Voor het verkrijgen van deze informatie heeft men instrumenten nodig. De bouwstenen voor deze instrumenten zijn items. In een itembank zijn een groot aantal items voorhanden, liefst meerdere items per doelstelling. Deze items zijn bovendien bij voorbaat bestudeerd naar diverse criteria. Met deze items kunnen, afhankelijk van de behoeften in het onderwijs, steeds nieuwe, desgewenst met elkaar vergelijkbare proeven samengesteld worden. Deze proeven duiden wij aan met de algemene benaming evaluatieproeven. Naargelang de proef als geheel en de items afzonderlijk aan bepaalde voorwaarden voldoen, of specifieke karakteristieken vertonen is de informatie die men m.b.v. deze proeven krijgt ook rijker te interpreteren. Men kan de algemene benaming van de proef dan toelichten met betekenisvolle bijvoeglijke naamwoorden. De proeven die het C.P.D.O. op het oog heeft kunnen aangeduid worden als *objectieve proeven voor didactische rendementsevaluatie*. Elk deel uit deze nogal lange benaming dekt een specifieke betekenis die wij kort zullen toelichten<sup>6</sup>.

De term *objectief* wijst op de uiterste zorg die tijdens elke constructiefase aan de proef besteed wordt om elke vorm van subjectiviteit uit te schakelen. Deze term heeft dan ook betrekking op de items afzonderlijk en op de proef als geheel. De items zijn instrumentele realiseringen van operabele leerdoelstellingen, de formulering is eenduidig zodat elke leerling, die de leerdoelstelling bereikt heeft, door geen enkel deel van het item misleid wordt. De quotering van de antwoorden gebeurt volgens vooraf opgestelde

instructies, zodat twee of meer beoordelaars noodzakelijk tot dezelfde uitslag komen. De proef wordt afgenomen volgens bepaalde richtlijnen. De gehele proef is een aselechte steekproef uit het universum van de leerdoelstellingen. Samenvattend kan men de objectieve proef omschrijven als een aantal eenduidig interpreteerbare, volstrekt objectief quoteerbare, doelstelling-valide items die representatief zijn voor het geheel van leerdoelstellingen uit een bepaald domein van inzichten, kennis en vaardigheden in een bepaald leervak, verworven gedurende een bepaalde onderwijsperiode.

Door *proef* wordt het verschil aangeduid enerzijds met de gewone schoolse proefwerken en anderzijds met tests, c.q. schoolvorderingentests. De proeven (in Nederland spreekt men meestal van schooltoets of toets) zijn evenwel meer verwant met een test dan met een proefwedstrijd, al zijn ook hier essentiële verschillen aan te geven.

Ten overstaan van de proefwerken onderscheidt de objectieve proef zich door inhoudsvaliditeit, doelstelling-validiteit, objectieve quoteerbaarheid, en standaardisatie, waardoor de vergelijkbaarheid van de verkregen resultaten gemaximaliseerd wordt. Ten overstaan van de schoolvorderingentest onderscheidt een objectieve proef zich hoofdzakelijk op drie essentiële punten: het doel dat men nastreeft, de methodiek die toegepast wordt bij de constructie en de studie van de metrische kwaliteiten.

Wat het doel betreft gaat de aandacht bij een schoolvorderingentest hoofdzakelijk uit naar het meten van het relatief peil dat een individuele leerling bereikt in de populatie waartoe hij behoort. Met een evaluatieproef tracht men van een individu, een groep of groepen het rendement (d.w.z. de mate waarin de doelstellingen gerealiseerd zijn) vast te stellen en vergelijkt men het bereikte resultaat met een a priori opgestelde norm.

Wat de constructie betreft, zowel van de items, als van de proef als geheel, verloopt deze voor de objectieve proef totaal op basis van een analytische lijst van leerdoelstellingen wat meestal niet geldt voor schoolvorderingentests.

Wat tenslotte de studie van de metrische eigenschappen betreft: deze vindt plaats bij de S.V.T. a posteriori. Bij een evaluatieproef gebeurt deze studie niet alleen uit noodzaak maar ook grotendeels uit principe a priori, wat evenwel de mogelijkheid en zelfs de wenselijkheid van een statistische verificatie achteraf niet overbodig maakt.

De term *didactisch* duidt de inhoud aan waarover men informatie wenst te verkrijgen, nl. de gedragsveranderingen, die als resultaat van de didactische arbeid op school nagestreefd worden in een bepaald domein van kennis en vaardigheden met de inhoud van een bepaald leervak dat gedurende een bepaalde periode onderwezen werd. De beoogde gedragsveranderingen werden vooraf geformuleerd in operabele doelstellingen die het didactisch handelen leiden bij de strategie-ontwikkeling. De leerdoelstellingen zijn dan ook de interne validiteitscriteria van elk item afzonderlijk en van de proef als geheel.

De *resultaten* die verkregen worden met een proef zijn niet direct interpreteerbaar. Om de waarde van een verkregen uitslag te interpreteren heeft men behoefte aan een vergelijkingsbasis. Men heeft twee mogelijkheden bij het kiezen van een vergelijkingsbasis. Of men vergelijkt de individuele uitslag met de statistische resultaten van een groep, die representatief is voor de populatie waartoe de leerling in kwestie behoort. Dit is een comparatief standpunt. Men kan dit comparatieve standpunt aanduiden met de term *peil* van het verkregen resultaat. Een andere vergelijkingsbasis is dat het verkregen resultaat beoordeeld wordt in het licht van een a priori opgestelde norm. Dit is een normatief standpunt, dat wij aanduiden met de term *rendement*. Met rendement wordt dan aangeduid de mate waarin het verkregen resultaat beantwoordt aan de eisen die men a priori heeft gesteld. Gezien de belangrijkheid komen wij in een volgend punt nog uitvoeriger op het probleem van de norm terug.

De term *evaluatie* refereert naar elk van beide interpretaties die men aan evaluatie toekent, nl. het evaluatieproces en de evaluatie-act. De proef is bedoeld om informatie te krijgen over

de mate waarin de leerdoelstellingen gerealiseerd zijn door de leerlingen als groep. De resultaten van de leerlingen moeten na verwerking geïnterpreteerd worden in het licht van de a priori gestelde norm. Voor deze interpretatie (evaluatie-act) is de informatie verkregen door de antwoorden van de leerlingen op de proef, een noodzakelijke maar nog onvoldoende basis om het onderwijsrendement te evalueren. Men heeft bovendien een tweede soort informatieve gegevens nodig, nl. die, welke de situationele realiteit beschrijven waarin de school haar taak moet vervullen. Deze gegevens hebben betrekking op descriptieve informatie over de leerlingen, de leerkrachten, de materiële en organisatorische modaliteiten waarin de school werkt.

Bovenstaande gegevens samenvattend is een objectieve proef voor didactische rendementsevaluatie te karakteriseren door de volgende vijf kenmerken:

- De proef als geheel is een representatieve, aselechte steekproef uit het universum van leerdoelstellingen uit een omschreven domein van kennis, inzichten en/of vaardigheden. De representativiteit heeft betrekking zowel op de leerinhouden als op de gedragingen die men met het onderwijs nastreeft.
- Elk item uit deze proef is een operationele definitie van een onderwijs-leersegment, dat voorafgaand operabel geformuleerd is in een leerdoelstelling.
- Elk item is bij voorbaat bestudeerd naar de congruentie met de leerdoelstelling die men met het item wil bevragen.
- Elk item is bij voorbaat bestudeerd naar de communicatieve perfectie, d.w.z. naar de eenduidige interpreteerbaarheid door de leerling.
- Elk item is bij voorbaat bestudeerd naar de objectieve scorebaarheid.
- Van elk item is bij voorbaat de normatieve (of theoretische) bereikbaarheid bekend.

### 2.1.2. Normatief standpunt als referentiebasis voor didactische evaluatie

Uit de resultaten die verkregen worden op een

objectieve proef volgt niet onmiddellijk een waardeoordeel over het onderwijsrendement. Hiertoe moeten de geobserveerde resultaten vergeleken worden met normen of criteria. Dit veronderstelt in de eerste plaats een stellingname t.o.v. de aard van de norm en vervolgens het bepalen van een methode om de norm vast te stellen.

De term *norm* wordt zowel in het algemeen spraakgebruik als in de wetenschappelijke research in twee betekenissen gebruikt, nl. in een statistische en in een ethische betekenis. In de psychometrische literatuur wordt meestal de statistische interpretatie gebruikt. 'Norm' betekent dan datgene wat men vaststelt in een populatie. De engelstalige term is 'normative measures'. In het Nederlands spreekt men van 'relatieve norm'<sup>7</sup> of 'comparatieve norm'<sup>8</sup>.

'Test norms may be defined as estimates of some characteristics of a distribution of test scores for a specified population. Norms describe the actual performance of a specified group of individuals'<sup>9</sup>.

De ethische betekenis legt de nadruk op datgene wat zou moeten zijn. In de engelstalige literatuur gebruikt men de term 'standards' of actueel, vooral onder invloed van de studies van Glaser, 'criterion-referenced measures'.

'Standards are desirable, or desired levels of attainment, preferably expressed in outcomes of instruction'<sup>10</sup>.

'Criterion-referenced measures depend on absolute standard of quality'<sup>11</sup>.

In het Nederlands wordt de ethische betekenis van norm aangeduid met 'absolute norm'<sup>12</sup> of 'normatieve norm'<sup>13</sup>. Wij willen erop wijzen dat de nederlandstalige term 'normatief' niet analoog is aan de engelstalige term 'normative'.

In de didactische literatuur wordt het comparatief standpunt voor evaluatieve doeleinden meer en meer verlaten ten gunste van het *normatief standpunt*<sup>14</sup>.

Wanneer men de prestaties van de leerlingen op een gegeven moment als norm van het onderwijs neemt, dan maakt men de norm afhankelijk van een feitelijke toestand die misschien,



zelfs waarschijnlijk, niet helemaal beantwoordt aan wat men van 'goed' onderwijs mag verwachten. Men waardeert integendeel de actuele prestatie in afhankelijkheid van de prestaties van de anderen of, zoals Ebel het uitdrukt 'relative working permits the student rather than the teacher to set the standards'<sup>15</sup>. Bovendien loopt men het gevaar, zoals Stellwag opmerkt, dat men het onderwijs bevriest<sup>16</sup>. Deze opmerking is evenwel niet louter toepasselijk op de empirische norm maar kan evenzeer betrekking hebben op het normatief criterium indien deze als 'onveranderlijk' beschouwd wordt. In dit opzicht wenst het C.P.D.O. de term 'absolute norm' te vermijden en stelt voor de term 'rendementscriterium' te gebruiken<sup>17</sup>.

Het zou oneerlijk zijn het standpunt van de voorstanders van het relatief standpunt niet te vermelden. Wijnen vraagt zich af of het wel mogelijk is a priori een norm te bepalen. Hij verwijst naar studies waarin aan onderwijspractici gevraagd wordt de bereikbaarheid van items te bepalen. Het resultaat dat met deze studies verkregen wordt is, volgens Wijnen niet zo hoopgevend<sup>18</sup>. Een tweede argument tegen het normatief standpunt is de vaststelling die de auteur zelf deed, nl. dat de bereikbaarheid van identieke vragen, in opeenvolgende jaren gesteld, niet constant is<sup>19</sup>.

Wat het eerste argument betreft willen wij opmerken dat empirische resultaten afhankelijk zijn van de waarde van het onderzoek en deze waarde is nauw verbonden aan de instrumentele realisering van de begrippen die men onderzoekt. De studies die Wijnen vermeldt laten in dit opzicht wel wat te wensen over. Zijn tweede argument pleit o.i. niet tegen de mogelijkheid of de wenselijkheid een rendementscriterium te stellen, integendeel. Het C.P.D.O. opteert voor het a priori stellen van een rendementsnorm uitgaande van de volgende gedachtengang. Het onderwijs wordt gegeven met het doel bepaalde gedragsveranderingen te verwezenlijken bij de leerlingen. Deze wenselijk geachte gedragsveranderingen worden operabel gedefinieerd in leerdoelstellingen. De leerdoelstellingen worden geïtemiseerd. Uit het resultaat op deze items zal

blijken of de beoogde doelstellingen door de leerlingen verworven zijn.

Deze normen die men stelt moeten aan de volgende vereisten voldoen: *a.* ze moeten een *realistische* verwachting uitdrukken, rekening houdend met de beperkende omstandigheid waarin de school in het algemeen werkt; *b.* anderzijds moeten ze de *theoretisch bereikbare* rendementsnorm aangeven, d.w.z. het hoogst mogelijke niveau aanduiden dat bij de betrokken populatie van leerlingen, bij optimaal onderwijs, haalbaar is<sup>20</sup>.

Wat de methode betreft om deze rendementscriteria te bepalen denken wij aan de beoordeling van de items door deskundige onderwijspractici<sup>21</sup>.

De *norm* die als vergelijkingsbasis genomen wordt *om de verkregen resultaten te interpreteren* is afhankelijk van het doel dat men met de proef nastreeft. Proeven voor selectiedoeleinden zullen andere normen aannemen dan evaluatieproeven. Evaluatieproeven die het C.P.D.O. op het oog heeft zijn in de eerste plaats bestemd om informatie te krijgen over de mate waarin *groepen leerlingen* (een klas, school, alle scholen van een provincie, enz.) de doelstellingen verworven hebben, waarvan de pedagogisch-didactische leiding verwacht dat ze nagestreefd worden in het onderwijs. Wij menen dat het een realistisch standpunt is aan te nemen, dat een doelstelling als bereikt beschouwd kan worden, wanneer in een heterogene groep m.b.t. de verstandelijke begaafdheid, de betere en de middelmatige leerlingen doelstelling-valide items juist beantwoorden. Naar analogie met de klassieke distributie van de intelligentie nemen wij aan dat de betere en de middelmatige leerlingen samen 75% van de bedoelde populatie uitmaken. Daaruit volgt dat wij als rendementscriterium een theoretische bereikbaarheid van 75% aannemen. Er wordt dan ook naar gestreefd bij het itemiseren van de leerdoelstellingen vragen te redigeren die een theoretische bereikbaarheid hebben van 75%.

Dit streven kan evenwel niet altijd gerealiseerd worden. Wij omschrijven de itemredactie immers als een instrumentale realisering van operabele doelstellingen. De bereikbaarheid

van een item wordt dan ook beïnvloed, ten eerste door de doelstelling als dusdanig en vervolgens door de graad van enkelvoudigheid waarin de leerdoelstelling uitgedrukt is. Hoe enkelvoudiger of hoe analytischer een leerdoelstelling geformuleerd is, hoe nauwer de relatie is tussen item en doelstelling. Hoe algemener of hoe complexer, hoe meer mogelijkheden er zijn, om uitgaande van eenzelfde doelstelling valide items in verschillende graden van bereikbaarheid te instrumentaliseren. Onze richtlijnen voor de bereikbaarheid in verband met itemiseren van leerdoelstellingen hebben dan ook hoofdzakelijk betrekking op deze complexere leerdoelen.

De redactiefase wordt verder gevolgd door de beoordeling van de items door een grote groep practici die, vanuit diverse beroepsbetrokkenheid met het onderwijs vertrouwd zijn<sup>22</sup>. Wij hopen uit de oordelen van deze rechters het rendementscriterium per item te kunnen bepalen. Dit laatste betekent evenwel niet dat empirische gegevens over bereikbaarheid overbodig zouden zijn bij het bepalen van de norm. Integendeel, in een volgende fase moeten de uitspraken van de experts op hun houdbaarheid getoetst worden. Hiertoe zullen de items die vooraf naar hun theoretische bereikbaarheid beoordeeld zijn, toegepast worden in klassen met een ongeselecteerde populatie, waarvan men weet dat het onderwijs er optimaal verzorgd wordt. Het gemiddelde resultaat dat door de leerlingen uit deze klassen bereikt wordt zou niet significant mogen afwijken van de beoordeling naar theoretische bereikbaarheid. Bovendien kunnen dezelfde items afgenomen worden in klassen met positieve selectie, d.w.z. klassen, samengesteld uit begaafde leerlingen met hoge studiemotivatie (voor het niveau einde basisonderwijs denken wij bijvoorbeeld aan sommige klassen van de zgn. voorbereidende afdeling tot het secundair onderwijs). Het gemiddelde resultaat van deze leerlingen zou hoger moeten zijn dan de beoordeling naar theoretische bereikbaarheid.

Voor zover de oordelen van de practici een realistisch normatief verwachtingspatroon weergeven, moeten de bereikte resultaten van ongeselecteerde groepen leerlingen, die optimaal

onderwijs hebben gekregen i.v.m. de betrokken doelstellingen, ongeveer gelijk zijn aan deze norm en moeten de resultaten van positief geselecteerde groepen zich boven deze norm situeren.

### 2.1.3 *Samenstellen van objectieve proeven*

Een grote voorraad vooraf bestudeerde items is als een magazijn waaruit items kunnen geput worden om steeds nieuwe, desgewenst met elkaar zeer nauw verwante, evaluatieproeven samen te stellen. Hierbij wordt verondersteld dat in de bank alleen items opgenomen zijn, waarvan na studie is gebleken dat ze perfecte instrumentaliseringen zijn van de betrokken leerdoelstellingen en dat deze items tevens eenduidig interpreteerbaar zijn. Wat de bereikbaarheid betreft kunnen o.i. zo mogelijk voor dezelfde leerdoelstelling items van verschillende graad van theoretische bereikbaarheid aanwezig zijn in de bank.

Naast de wenselijkheid voor complexere leerdoelstellingen, items van verschillende bereikbaarheid voorhanden te hebben, zijn er ook leerdoelstellingen die zo fundamenteel zijn, dat ze door nagenoeg alle leerlingen verworven moeten zijn en leerdoelstellingen die niet door 75% van een ongeselecteerde groep haalbaar zijn. Ook deze leerdoelstellingen moeten in geïtemiseerde vorm in de bank voorhanden zijn. In de mate dat men uit de itembank immers niet louter evaluatieproeven (al is dit de hoofdfunctie van een itembank) maar bijv. ook proeven voor selectie met het oog op verschillende richtingen van voortgezet onderwijs wil samenstellen, zijn deze items noodzakelijk in een itembank.

Voor het opnemen van items in evaluatieproeven denken wij, voorlopig althans, dat alleen items in aanmerking komen waarvan de theoretische bereikbaarheid 75% of meer bedraagt. De stelling, dat de bereikbaarheid hoger mag zijn dan 75% wijkt enigszins af van een vroegere stellingname, waar wij louter voor de 75% grens opteerden<sup>23</sup>. De reden van deze wijziging baseert zich op het beoordelingsonderzoek van de leerdoelstellingen. Er komen inderdaad een aantal leerdoelen voor in de inventaris die

volgens het oordeel van de practici door meer dan 75% van de leerlingen bereikt moeten zijn. Het zou onverantwoord zijn deze doelstellingen uit de evaluatieproef te weren.

Een analytische lijst van leerdoelstellingen overtreft in ruime mate het aantal items die redelijkerwijze in een proef opgenomen kunnen worden. De doelstellingen die in de proef opgenomen worden moeten evenwel een representatieve steekproef zijn uit het universum van de doelstellingen om de verkregen resultaten te kunnen generaliseren over het geheel van de leerdoelstellingen.

In handboeken voor statistiek en psychometrie worden verschillende methoden voor het nemen van steekproeven besproken. Bij evaluatieve doeleinden, waarbij men niet zozeer een individueel resultaat maar het resultaat van groepen op het oog heeft, wordt men met een meer complexe structuur van het nemen van steekproeven geconfronteerd. Men kan immers wensen te onderzoeken in welke mate *de* doelstellingen bij *de* leerlingen verworven zijn. Fundamentele studies hierover werden verricht door F. Lord, die deze benadering aanduidde met de term 'item sampling'<sup>24</sup>. De voorstellen van Lord werden verder uitgewerkt door Husek en Sirotnik die van 'matrix sampling' spreken<sup>25</sup>. De voorgestelde werkwijzen werden reeds toegepast o.a. door Cook en Stufflebeam<sup>26</sup>, Owens en Stufflebeam<sup>27</sup> en Husek en Sirotnik<sup>28</sup>.

De resultaten, in deze studies verkregen, zijn zeer positief voor de estimatie van de gemiddelden. Ze zijn evenwel minder positief voor de estimatie van de varianties.

Husek en Sirotnik onderscheiden zes mogelijke patronen voor het nemen van steekproeven<sup>29</sup>.

- a. Statistische inferentie of item-steekproeven. Een steekproef van items wordt afgenomen van de populatie leerlingen.
- b. Psychometrische inferentie of subject-steekproeven. Alle doelstellingen worden bevroegd bij een steekproef van de leerlingen.
- c. Enkelvoudige statistisch-psychometrische inferentie of enkelvoudige item- en subject-steekproeven. Een steekproef uit het univer-

- sum van items wordt afgenomen van een steekproef uit de populatie van de leerlingen.
- d. Statistisch-psychometrische inferentie met twee subject-steekproeven en één item-steekproef.
- e. Statistisch-psychometrische inferentie met het nemen van meervoudige itemsteekproeven en meervoudige subject-steekproeven.
- f. Statistisch-psychometrische inferentie met het nemen van meervoudige itemsteekproeven en meervoudige subject-steekproeven uit twee subject-populaties.

We kunnen in onderhavige bijdrage op deze voorstellen niet verder ingaan. Voor didactische evaluatie van het onderwijsrendement zijn de voorstellen van matrix-sampling evenwel belangrijk. Door toepassing van het meervoudig statistisch-psychometrisch inferentie-patroon kunnen meerdere, elkaar niet overlappende steekproeven uit de inventaris van leerdoelstellingen genomen worden, die aan steekproeven van leerlingen worden voorgelegd. Op deze wijze krijgt men informatie over een zeer groot aantal leerdoelstellingen terwijl elke steekproef van leerlingen slechts een redelijk aantal items moet beantwoorden. Bij verdere studie van de constructie van evaluatie-instrumenten moet deze matrix-sampling zeker verder bestudeerd worden.

Voor het samenstellen van evaluatieproeven denken wij hoofdzakelijk aan gestratificeerde steekproeven. Hiervoor is het nodig dat de itembank gestructureerd is, d.w.z. ingedeeld volgens bepaalde criteria. De criteria die hiervoor in aanmerking komen zijn mede afhankelijk van het gebruik dat men van de bank wil maken, dus van de aard en de diversiteit der diensten die men met de bank wil verlenen. De aan te brengen structuren kunnen variëren van enkelvoudig tot zeer complex. Men kan onderscheid maken tussen drie soorten gegevens:

- a. inhoudelijke gegevens (belangrijk voor inhoudelijk gestratificeerde steekproeven)
  - het leervak, rubriek, subrubriek, topic of leerdoelstelling waarop het item betrekking heeft;

- de cognitieve operatiecategorie waarvoor het item valide is;
- het leerjaar waarvoor het item in de eerste instantie geëigend is, eventueel aan te vullen met aanduiding van de leerjaren waarvoor het item eveneens dienstig kan zijn.

b. metrische gegevens

- aanduiding van het juiste antwoord op het item;
- het resultaat van de studie naar doelstelling-validiteit;
- het resultaat van de studie naar eenduidigheid;
- de normatieve of theoretische bereikbaarheid van het item;
- de verwachte of hypothetische feitelijke bereikbaarheid (geschat door een beoordelingsgroep);
- empirisch gevonden bereikbaarheid.

c. Gegevens m.b.t. instrumentaliseringvorm.

Deze gegevens kunnen bij elk item worden opgenomen. Voor de materiële vormgeving kan men denken aan randponskaarten of wanneer men over de mogelijkheid beschikt met een computer te werken dan kunnen de items genoteerd worden hetzij op ponskaarten of nog beter op de band. The Computer System Section of the Center for Educational Development of the University of Illinois ontwikkelde een systeem bekend onder de naam CRIB (Computerized Random Item Bank)<sup>30</sup>. De items zijn opgenomen op de band en deze wordt verbonden met een on-line writer. Het programma is zeer eenvoudig. De items zijn per topic gegroepeerd. Men kiest een bepaalde topic en de computer neemt een aselechte steekproef uit het universum van de voorraad items, volgens een in het programma opgenomen generator.

Een ander voorbeeld van het nemen van steekproeven uit een itembank met behulp van de computer, is het systeem dat gebruikt wordt in 'the Wayne State Medical School'<sup>31</sup>. Dit systeem is meer complex, in die zin dat naast het inhoudelijk aspect tevens rekening wordt ge-

houden met verschillende statistische parameters van het item.

2.1.4. *Evaluatieproeven verwant aan de rendementsproeven van het C.P.D.O.*

In de bovenstaande gegevens werd bijna uitsluitend ingegaan op de proeven voor didactische rendementsevaluatie die het C.P.D.O. op het oog heeft. Een soortgelijke problematiek wordt evenwel ook door andere onderzoekers en researchcentra bestudeerd. We bedoelen voornamelijk de criterion-referenced measures, de mastery-tests en de branch-tests of tailored tests. Deze proeven onderscheiden zich hoofdzakelijk op drie punten van de rendementsproeven van het C.P.D.O.

- a. Ze zijn in eerste instantie ontworpen voor individuele evaluatie.
- b. Ze gaan veel verder in de richting van de descriptieve diagnose dan de proeven aan het C.P.D.O.
- c. Ze bevatten geen rendementsnorm. De verkregen informatie geeft aan wat de leerling kent of nog niet kent, maar daarmee weet men nog niet of dit als een voldoende rendement beschouwd kan worden.

We zullen de essentiële karakteristieken van elk der drie vermelde proeven samenvattend weer-geven.

*Criterion-referenced measures*

Deze proeven worden sinds 1963 vooral bestudeerd in het kader van het project 'Individually Prescribed Instruction' in het 'Learning Research and Development Center of Pittsburgh'<sup>32</sup>. De term 'criterion-referenced test' betekent:

'a test that is deliberately constructed in terms of yield scores that are directly interpretable in terms of specified performance standards'<sup>33</sup>.

Deze proeven worden gekarakteriseerd door vier kenmerken: - de gedragscategorieën voor de verschillende graden van kennis zijn zo duidelijk mogelijk geformuleerd vóór de proef geconstrueerd wordt; - elke gedragscategorie is



in meerdere items geoperationaliseerd; – men neemt een aselechte steekproef uit de items; – het verkregen resultaat drukt de kennis of kunde uit voor de geïnventariseerde gedragscategorieën<sup>34</sup>.

De constructie van de criterion-referenced test verloopt in algemene fasen identiek aan de proeven die het C.P.D.O. op het oog heeft. Men gaat uit van een analytische lijst van 'a domain of instructional tasks' die omgezet worden in termen van waarneembare gedragingen. Voor elk van deze doelstellingen worden meerdere items geschreven. De karakteristiek van deze items is dat ze direct afgeleid worden van de doelstelling. Meerdere items per onderzochte doelstellingen worden aan de leerling voorgelegd en uit zijn antwoorden leidt men af of deze leerling de doelstelling al of niet verworven heeft<sup>35</sup>.

#### *Mastery test*

B. S. Bloom's opvattingen over evaluatie zijn zeer nauw verwant met het concept van criterion-referenced measures<sup>36</sup>.

Bloom gaat uit van de opvatting dat de basis-kennis en vaardigheden die elke mens nodig heeft om als volwaardige burger te kunnen leven door alle leerlingen (of beter door 95%) van de leerlingen moeten kunnen verworven worden. Sommigen zullen hier evenwel minder tijd voor nodig hebben dan anderen. Goed onderwijs moet daarom rekening houden met de individuele verschillen van de leerlingen en het resultaat dat verkregen wordt moet eveneens individueel benaderd worden. Voor ieder leersegment dat doorgewerkt is krijgt de leerling een masterytest. Deze bestaat uit meerdere opgaven voor elk der vooraf geformuleerde leerdoelstellingen. Uit de test moet de 'mastery' of 'nonmastery' blijken, Indien de leerling de doelstelling niet bereikt heeft, moet tevens uit de antwoorden van de leerlingen een gedetailleerde descriptieve diagnose gemaakt worden over wat de leerling nog moet leren vóór hij de doelstelling volledig zal beheersen. De basis waarop Bloom steunt voor zijn theorie is drievoudig: enerzijds op de leer-

theorie van Carroll<sup>37</sup> en anderzijds, voor het opstellen van de doelstellingen, op de theorie van Gagné over de hiërarchische structuur van de leerinhouden<sup>38</sup> en de 'taxonomy of educational objectives' die onder zijn leiding tot stand kwam<sup>39</sup>.

#### *Branch tests of tailored tests*

Branch-testing gaat in de richting van Bloom nog een stap verder. Elk volgend item dat een leerling zal beantwoorden, wordt bepaald door zijn antwoord op het (of de) vorige item(s)<sup>40</sup>

Deze procedure maakt het mogelijk dat elke leerling een individueel testpatroon volgt, zodat men uit zijn antwoorden een zeer duidelijk beeld heeft van de doelstellingen die de leerling in kwestie bereikt heeft en van die, welke hij nog niet bereikt heeft. Branch-testing wordt hoofdzakelijk toegepast voor diagnostische doeleinden. De ontwikkeling van deze tests is zeer recent en tot heden werd nog weinig onderzoek verricht over de mogelijkheden tot gebruik van branch-testing<sup>41</sup>.

#### *2.2. Bijkomende functies van een itembank*

Naast het voorhanden hebben van een grote voorraad goede items kan een itembank nuttige diensten bewijzen bij instructie en coördinatie.

##### *2.2.1. Mogelijke instruerende functie van een itembank*

Aangezien een itembank een groot aantal items bevat, die instrumentele realiseringen zijn van doelstellingen, kunnen deze items als hulpmiddel gebruikt worden bij instructiedoeleinden.

a. Men kan aan de hand van items op zeer concrete wijze de inhoud, de bedoeling en de draagwijdte van de leerdoelstellingen en van het leerplan illustreren en verduidelijken<sup>42</sup>.

b. Items kunnen een zeer suggestieve hulp zijn bij de voorbereiding van een les of lessenreeks. Een goede itembank geeft immers een gedetailleerd beeld van de verschillende gedrags-

vormen die de leerlingen met de leerstofinhoud moeten kunnen uitvoeren<sup>43</sup>.

c. Tal van vakdidactische principes kunnen aan de hand van items duidelijk gedemonstreerd worden.

### 2.2.2. Coördinerende functie van een itembank

De coördinerende functie van een itembank concretiseert zich in de volgende situaties.

a. Elk jaar wordt in tal van scholen (behorende tot een zelfde onderwijsniveau), onafhankelijk van elkaar, geëxamineerd. Dat betekent dat zoveel verschillende leerkrachten geregeld voor dezelfde problematiek staan, nl. op een verantwoorde wijze een oordeel uitspreken over de prestaties van de leerlingen. Voor een aanzienlijk gedeelte heeft deze oordeelsvorming plaats op basis van vragen over de doorgemaakte leerinhouden. Dit veronderstelt dat zoveel leerkrachten, onafhankelijk van elkaar, een aantal vragen moeten redigeren op basis waarvan ze tot de gewenste oordeelsvorming zullen komen. Deze werkwijze is noch economisch noch efficiënt. Een itembank kan hier zeer goede diensten bewijzen door items ter beschikking te stellen aan de scholen. Bovendien verdient een goed item een langer leven dan een eenmalig gebruik ergens voor enkelingen, of zoals Wood het nogal humoristisch uitdrukt '... good items are not like woman's fashions; they are inherently durable and at the usual rate of syllabus change, many can survive for a number of years'<sup>44</sup>.

b. In de traditionele werkwijze redigeert elke individuele leerkracht de vragen voor zijn leerlingengroep. Men begrijpt evenwel dat de interindividuele verschillen in bekwaamheid voor het redigeren van vragen tussen de leerkrachten aanzienlijk kan zijn. De kwaliteit van de gestelde vragen zal dan eveneens in evenredige mate van elkaar verschillen. Een itembank, waarvoor men een beroep kan doen op de medewerking van meerdere uitgelezen en getrainde itemredactoren, kan hier nuttige dienst verlenen<sup>45</sup>.

c. De vragen worden eveneens beoordeeld en gequoteerd door de individuele leerkracht. Daarin zitten natuurlijk een aantal positieve en verdedigbare aspecten, doch de leerkracht heeft louter een subjectieve vorm (wat daarom nog niet pejoratief geïnterpreteerd moet worden). Meer objectieve normen kunnen in een itembank worden voorzien en op deze wijze coördinerend werken tussen de verschillende scholen.

### 3. Enkele voorlopers en actuele projecten van itebanking

Noch het idee, noch de uitvoering voor het aanleggen van een voorraad 'goede' items zijn helemaal nieuw in de strikte zin van het woord. Er kunnen sporen aangeduid worden die erop wijzen dat meer dan 20 eeuwen voor onze tijdrekening in China een rudimentaire verzameling vragen bestond. Gedurende de Khandynastie (rond 1100 vóór Christus) groeide deze embryonale itembank zelfs uit tot min of meer georganiseerde verzamelingen van vragen over muziek, wiskunde, spelling, sterrenkunde en kennis en vaardigheden op gebied van paardrijden<sup>46</sup>.

Het kan moeilijk de bedoeling zijn in deze bijdrage een volledige beschrijving te geven van de historische evolutie i.v.m. itebanking. Toch is het nuttig enkele hoofdmomenten uit de geschiedenis even te herdenken. In dit verband moet gewezen worden op de gedachte en het werk van een tweetal personen in de tweede helft van de negentiende eeuw. In de eerste plaats betreft het de ideeën en het werk van Horace Mann<sup>47</sup>. In een analytische studie over de mogelijkheden en de grenzen van examens verdedigt deze auteur op pertinente wijze de schriftelijk te beantwoorden vragen. Bovendien stelde hij in Boston een interscolair proefwerk op voor de public schools. Deze examens testten de kennis van de leerlingen in verband met geschiedenis, aardrijkskunde, natuurwetenschappen, wiskunde, spraakkunst en astronomie. Ten behoeve van deze proeven werden vragen opgesteld, en beschikte men over een zeer bescheiden pool die

voor volgende jaren bewaard werd<sup>48</sup>.

Waarschijnlijk nog explicieter komt deze bezorgdheid voor het aanleggen van een item-pool tot uiting in het zgn. 'Scale Book' van George Fisher. Deze Engelse onderwijspracticus poogde niet louter items te bewaren, maar had reeds realisaties in de richting van duidelijke structuren en elementen die de eigenschappen der items aangeven. De items uit het Scale Book waren bedoeld als voorbeelden zoals blijkt uit onderstaand citaat.

'Questions (of several branches) are contained in the Scale Book, to serve as type, not only of the difficulty, but of the nature of the question'<sup>49</sup>.

Ten overstaan van het werk van Fisher merkte Ayres op dat deze auteur zijn tijd evenwel zo ver vooruit was dat zijn ideeën niet aansloegen bij zijn collega's, tenminste te oordelen naar de geringe weerklink die zijn werk gekend heeft<sup>50</sup>.

Vanaf het einde van de vorige eeuw en het begin van deze eeuw begint de wetenschappelijke studie van de leerresultaten met de studies van Rice<sup>51</sup> en Stone<sup>52</sup>. Deze eerste benaderingen accentueerden hoofdzakelijk het aspect van objectieve meting. Beide studies hadden op zichzelf alleen onrechtstreeks te maken met de ontwikkeling van itembanken. We noemen het onrechtstreeks omdat geen van deze studies op zichzelf het probleem stelde, maar wel in de verdere evolutie van de 'achievement testbeweging' hoofdzakelijk wetenschappelijk-commerciële instellingen voorraden goede items gingen aanleggen.

De specifieke betekenis, die de term itembank actueel heeft, m.n. een gestructureerde voorraad van items die doelstelling-valide zijn, dateert van zeer recente datum. De eerste publikatie die het probleem besprak is het verslag van Wood in 1967<sup>53</sup>. De auteur publiceerde een interim-rapport van een studie, ondernomen door de NFER (National Foundation for Educational Research). Het doel van deze studie bestond in een zgn. pilot study naar de mogelijkheden van het aanleggen van een itembank voor CSE-examination (Certificate of Secondary Education). De ontwikkeling van de itembanken is zeer nauw verbonden aan de evaluatiebeweging. Evaluatie

werd immers door Tyler in de dertiger jaren gedefinieerd als het nagaan in welke mate de doelstellingen van het onderwijs gerealiseerd zijn bij de leerlingen<sup>54</sup>. Dit veronderstelde 'the development of instruments appropriate to the newly created needs'<sup>55</sup>. Eveneens daterend van het eind der jaren dertig werd dezelfde bedenking teruggevonden in de dossiers van de Harcourt Publishing Test Company, '...warning that this movement was coming, and urging the company to prepare for it'<sup>56</sup>. Ondanks deze expliciete bedenkingen moest men evenwel tot 1965 à 1970 wachten voor, eerst in wetenschappelijke researchdiensten en later in wetenschappelijk-commerciële test-organismen, een begin gemaakt werd met de itemconstructie gebaseerd op leerdoelstellingen. Actueel zijn meerdere projecten in ontwikkeling, waarvan voorlopig uitsluitend interim-rapporten voorhanden zijn omdat de studies amper begonnen zijn. Naast het NFER-project in Engeland<sup>57</sup> is het best bekende banking-project waarschijnlijk dat van Popham van het Center for the Study of Evaluation van de University of California in Los Angeles. Popham organiseert een 'Instructional Objectives Exchange' (IOE). Het doel dat hij zich stelt is het verzamelen en organiseren van relevante leerdoelstellingen en deze te laten samengaan met items die deze leerdoelstellingen vertegenwoordigen<sup>58</sup>. Een ander, reeds ruim ontwikkeld project is het EII (Evaluation for Individualized Instruction). Stafleden van het Institute for Educational Research van de University of Illinois, trainden leraren voor het redigeren van leerdoelstellingen en items. In 1972 beschikte dit project over 5000 leerdoelen en 25.000 items<sup>59</sup>. Eveneens in de 'University of Illinois' bestaat een tweede project, bestudeerd door CED (Center for Educational Development)<sup>60</sup>.

In Duitsland lopen een tweetal projecten. Het eerste is het Itembank-project van de Pädagogische Hochschule in Kiel, het tweede is dat van de 'Landesregierung' van Schleswig-Holstein. Dit laatst genoemde project begon in 1971 en loopt tot 1976<sup>61</sup>.

In Nederland worden elk jaar door het Centraal Instituut voor Toetsontwikkeling

(C.I.T.O.), in samenwerking met diverse werkgroepen leerkrachten honderden items geconstrueerd. Een gedeelte van deze items wordt opgenomen in de 'schooltoetsen'. Wij kunnen de verzameling van deze items evenwel geen didactische itembank noemen in de strikte zin van het woord, daar de items niet opgesteld zijn op basis van geëxpliceerde leerdoelstellingen.

Wetenschappelijk-commerciële organisaties zoals ETS (Educational Testing Service) en Harcourt Brace Jovanivick Incorporation begonnen pas in 1971 met itembanking voor 'behavioural stated objectives'<sup>62</sup>.

Bovenstaande gegevens duiden exemplarisch enkele actuele projecten aan die itembanking bestuderen. Geen van deze studies is voltooid, alle zijn nog in volle ontwikkeling.

#### Noten

1. E. de Corte, *Onderwijsdoelstellingen. Bijdrage tot de didaxologische theorievorming en aanzetten voor empirisch onderzoek*. (Studia Paedagogica, nr. 2), 1973, IX-200.
2. L. J. Cronbach, Course improvement through evaluation. *Teachers College Record*, 1963 (64), p. 678;
  - M. Scriven, The methodology of evaluation. In: R. W. Tyler, R. M. Gagne & M. Scriven, *Perspectives of curriculum evaluation*. (AERA-monograph series on curriculum evaluation, nr. 1), 1967, p. 56-58;
  - R. Wood & L. S. Skurnik, *Item banking. A method for producing school-based examinations and nationally comparable grades*, 1969, p. 1;
  - B. S. Bloom, J. T. Hastings & G. F. Madaus, *Handbook on formative and summative evaluation of student learning*. 1971, p. 38 & 251-253;
  - W. Royl, Itembank: Aufbau und Funktion. *Die Deutsche Schule*, 1972 (64), nr. 9, 547-564.
3. B. S. Bloom, J. T. Hastings & G. F. Madaus, *o.c.*, p. 38 & p. 251;
  - R. L. Thorndike, Educational measurement for the seventies. In: R. L. Thorndike (Ed.), *Educational measurement*. 1971, p. 6;
  - W. Royl, *o.c.*, p. 547.
4. B. S. Bloom geeft een soortgelijke omschrijving van itempool. Zijn omschrijving is evenwel enger dan de onze. Wij spreken immers van een structurering maar bepalen bij voorbaat de structuren niet noch specificeren welke parameters bekend moeten zijn. Verder menen wij, in tegenstelling met Bloom, dat een itembank functioneel breder is dan het samenstellen van evaluatieproeven. Cf. B. S. Bloom, J. T. Hastings & G. F. Madaus, *o.c.*, 1971, p. 251.
5. In een volgende bijdrage hopen wij verder in te gaan op enkele problemen verbonden aan de opbouw van een didactische itembank, inzonderheid op de redactie van items als instrumentele realisering van operabele doelstellingen.
6. In een vroegere publikatie heeft het C.P.D.O. de theoretische principes geformuleerd i.v.m. de evaluatie van het pedagogisch-didactisch rendement van het onderwijs. In de volgende uiteenzetting zal nu en dan gebruik gemaakt worden van de gegevens uit deze publikatie. Cf. K. Swinnen, E. de Corte & G. Tistaert, Een theoretisch kader voor studie en onderzoek betreffende de evaluatie van het pedagogisch-didactisch rendement van het onderwijs. *Tijdschr. Opvoedk.*, 1968 (13), nr. 3, 129-171.
7. W. H. F. Wijnen, *Onder of boven de maat. Een methode voor het bepalen van de grens voldoende/onvoldoende bij studietoetsen*. 1971, p. 22.
8. K. Swinnen, E. de Corte & G. Tistaert, *o.c.*, p. 147.
9. J. C. Flanagan, Units, scores and norms. In: E. F. Lindquist (Ed.), *Educational measurement*. 1963, p. 698.
10. *Ibidem*, p. 698.
11. R. Glaser, Instructional technology and the measurement of learning outcomes. Some questions. *American Psychologist*, 1963 (18), p. 519.
12. W. H. F. Wijnen, *o.c.*, p. 21.
13. K. Swinnen, E. de Corte & G. Tistaert, *o.c.*, p. 147.
14. R. Glaser & A. J. Nitko, Measurement in learning and instruction. In: R. L. Thorndike (Ed.), *Educational measurement*. 1971 (2nd ed.), p. 653;
  - Dorothy G. Mueller, How to evaluate teaching. *J. Teacher Educ.*, 1971 (22), nr. 2, p. 232;
  - M. Roebuck, Floundering among measurements in educational technology. *Programmed learning and educational technology*, 1972 (9), nr. 2, p. 93;
  - J. O. Vittetoe, Evaluation of product. An essential first step. *J. Teacher Educ.*, 1972 (23), nr. 2, p. 130.



15. R. L. Ebel, *Measuring educational achievement*. 1965, p. 408.
16. Helena W. F. Stellwag, Schoolvorderingentest als selectiemiddel. In: *Voor- en nadelen van de schoolvorderingentest*. 1966, p. 18.
17. E. de Corte, *Didactische evaluatie en doelstellingen in het onderwijs. Bijdrage tot de wetenschappelijke theorievorming en ontwikkeling van onderzoeksmethoden*. 1970, p. 151.
18. W. H. F. Wijnen, *o.c.*, p. 27.
19. W. H. F. Wijnen, Studietoets: beoordeling van studie en onderwijs. In: R. R. Gras & R. A. P. Tielmans (Eds), *Tweede nationaal congres, onderzoek van wetenschappelijk onderwijs*. 1968.
20. E. de Corte, *o.c.*, p. 152.
21. G. Tistaert, *Instrumentalisering van leerdoelstellingen in functie van didactische rendements-evaluatie. Ontwerp van een methode voor redactie en beoordeling van meerkeuzevragen met inbegrip van empirisch onderzoek*. 1973, p. 219 e.v.
22. *Ibidem*, p. 217-233.
23. K. Swinnen, E. de Corte & G. Tistaert, *o.c.*, p. 141.
24. F. M. Lord, *Item sampling in test theory and research design*. 1965, 27 pp.
25. T. R. Husek & K. Sirotnik, *Item sampling in educational research*. C.S.E.I.P. Occasional Report, nr. 2, 1967, 32 pp.
26. D. L. Cook & D. L. Stufflebeam, Estimating test norms from variable size item and examinee samples. *Educ. Psychol. Measmt*, 1967 (27), 601-610;  
- D. L. Cook & D. L. Stufflebeam, Estimating test norms from variable size item and examinee samples, *J. Educ. Measmt*, 1967 (4), 27-33.
27. T. R. Owens & D. L. Stufflebeam, *An experimental comparison of item sampling and examinee sampling for estimating test norms*. Paper read at the annual meeting of the AERA 1968, 14 pp.
28. T. R. Husek & K. Sirotnik, *Matrix sampling in educational research: an empirical investigation*. Paper read at the annual meeting of the AERA, 1968, 15 pp.
29. T. R. Husek & K. Sirotnik, *o.c.*, p. 7-19.
30. J. J. Marxer, Computer storage and retrieval of test items. In: H. A. Curtis (Ed.), *The development and management of banks of performance based test items*. 1972, p. 33-34.
31. *Ibidem*, p. 35.
32. R. Glaser, Instructional technology and the measurement of learning outcomes. *American Psychologist*, 1963 (18), 519-521;  
- R. Glaser & R. C. Cox, Criterion-referenced testing for the measurement of educational outcomes. In: R. Weisgeber (Ed.), *Instructional process and media innovation*. 1968, 545-550;  
- A. J. Nitko, *Criterion referenced testing in the context of instruction*. Paper presented at the N.C.M.E., 1970, 19 pp.
33. R. Glaser & A. J. Nitko, Measurement in learning and instruction. In: R. L. Thorndike (Ed.), *Educational measurement*. 1971 (2nd ed.), p. 653.
34. A. J. Nitko, *A model for criterion-referenced tests based on use*. Paper read at the AERA, 1971, p. 2.
35. J. H. Block, Criterion-referenced measurements: potential. *School Rev.*, 1971 (79), p. 290.
36. B. S. Bloom, J. Th. Hastings & G. F. Madaus, *Handbook on formative and summative evaluation of student learning*. 1971, cf. 43-57.
37. J. A. Carroll, A model of school learning. *Teachers College Record*, 1963 (64), 723-733.
38. R. M. Gagné, *The conditions of learning*. 1965.
39. B. S. Bloom (Ed.), *Taxonomy of educational objectives. The classification of educational goals. Handbook I: Cognitive domain*. 1956.
40. R. Glaser & Nitko, *o.c.*, p. 651-652.
41. *Ibidem*, p. 651.
42. W. J. Popham, *Objectives and instructions*. Paper read at the AERA meeting, 1969 p. 3.
43. W. Royl, Itembank. Aufbau und Funktion. *Die Deutsche Schule*, 1972 (64), nr. 9, p. 560-561.
44. R. Wood & L. S. Skurnik, *Item banking. A method for producing school based examinations and nationally comparable grades*. 1969, p. 3.
45. W. Royl *o.c.*, p. 559.
46. P. H. Dubois, A test dominated society: China, 115 BC 1905 AD. In: *Proceedings of the 1964 invitational conferend on testing problems*. 1965, p. 5.
47. H. Mann, Report of the annual examining committees of the Boston grammar and writing schools. *Common School J.*, 1845 (21), nr. 7, 326-336. Gecit. in: J. C. Stanley, *Measurements in today's school*. 1965, p. 12.
48. O. W. Caldwell & S. A. Courtis, *Then and now in education. 1845-1923*. 1925, p. 4. Gecit. in: V. H. Noll, *Introduction to educational measurement*. 1957, p. 19.
49. From a letter quoted in E. Chadwick, *Statistics of educational results*. Geciteerd in: J. C.

- Stanley, *Measurement in today's school*. 1965, p. 13.
50. L. P. Ayres, History and presents notes of educational measurements. In: *The measurement of educational products. Seventeenth Yearbook of the NSSSE*. Bloomington, 1918, p. 10.
51. J. M. Rice, The futility of the spelling grind. *Forum*, 1897 (23), 163-172, 409-419.
52. C. W. Stone, *Arithmetical abilities and some factors determining them*. 1908.
53. R. Wood, *The itembank project: a pilot study of an alternative method of calibrating attainment in the CSE examination*. Interim research bulletin, 1967.
54. R. W. Tyler, *Constructing achievement test*. 1934, p. 18.
55. D. E. Scates, The improvement of classroom testing. *Rev. Educ. Res.*, 1938 (8), p. 523.
56. H. A. Curtis, Commercially produced item banks. The local project director's responsibilities. In: H. A. Curtis (Ed.), *o.c.*, 1972, p. 12.
57. R. Wood & L. S. Skurnik, *o.c.*, 1969.
58. W. J. Popham, *Objectives and instruction*. Paper read at the annual meeting of the AERA, 1969, p. 3;  
- W. J. Popham, Objectives based management strategies for large educational systems. *J. Educ. Res.*, 1972 (66), nr. 1, p. 7.
59. T. J. Slocum, Locally produced itembanks. In: H. A. Curtis (Ed.), *o.c.*, 1972, p. 7.
60. *Ibidem*, p. 7-8.
61. W. Royl, Itembank: Aufbau und Funktion. *Die Deutsche Schule*, 1972 (64), nr. 9), p. 558.
62. Muriel M. Abbott, Publisher's management problems when entering into a new field of test development. In: H. A. Curtis (Ed.), *o.c.*, 1972, p. 19.

#### Curriculum vitae

G. Tistaert, geboren te Asse in 1932, studeerde Pedagogische Wetenschappen aan de Katholieke Universiteit te Leuven, waar hij in 1961 het licentiaatsdiploma behaalde en in 1973 promoveerde tot doctor.

Gedurende een tiental jaren was hij als pedagogiek leraar werkzaam in een Normalschool voor de opleiding van toekomstige onderwijzers. In 1964 werd hij wetenschappelijk medewerker aan het C.P.D.O. van de Leuvense Universiteit. Thans is hij als docent aan dezelfde universiteit verbonden waar hij zich hoofdzakelijk bezig houdt met de opleiding van toekomstige pedagogiek leraren.

Adres: Katholieke Universiteit, Dekenstraat 28-30, Leuven (België).