

[toetsvragen ontwerpen / schrijven]

vraagconstructie 1

ben wilbrink 8-3-76.

Shoemaker, D. M. Toward a framework for achievement testing.
Review of Educational Research 1975, 45, 127147

"Everyone knows that an achievement test must be related to the instructional program, but the nature of this relationship has always been ambiguous or mysterious, and the resultant problems have become unbearable. So sad is the current state of affairs that we can barely distinguish between a 'good' and a 'bad' achievement test and must rely upon expert opinion when the difference should be obvious to every student in the instructional program."

De oplossingen die door Shoemaker voorgestaan worden behelzen constructie van een vragenverzameling die isomorf is aan het onderwijsprogram (of omgekeerd, zoals je wilt). Wat ons hier interesseert is niet zozeer de problematiek rond het aanleggen en gebruiken van een dergelijke verzameling (dat komt in andere onderdelen van het moduul toets constructie aan de orde) maar in de wijze waarop vragen geconstrueerd worden. Daar onderscheidt Shoemaker drie mogelijkheden die op het ogenblik in ontwikkeling zijn: de itembanking approach, de itemforms approach,, en de transformationalgrammar approach..

De itembanking approach. Daar hebben we voor onze vraagstelling niets aan, omdat het probleem van constructie van de vragen zelf overgelaten wordt aan 'item constructeurs' die het dan zelf maar op hun traditionele wijze aan moeten pakken.

De itemforms approach.

"In the itemforms approach, the goals of an instructional program are reduced to statements of highly specific objectives, and, for each specific objective, one or more item forms are constructed. Here, the itemdomain consists of all items capable of being generated by itemforms. An itemform is a rule or algorithm for generating a class of specific test items (onderstreping van

mij). To demonstrate what is meant here, consider the following specific objective taken from elementary arithmetic: 'Students will be able to compute the sum of any two singledigitpositive nonzero integers presented in horizontal format.' For this objective, the corresponding item form might be

$$(a_1 + a_2 = --$$

where the possible values for $a(i)$ are taken from the set (1,2,3,4,5,6,7,8,9). The itemform defines clearly and unambiguously a set of items where each item is capable of generating 81 specific test items where each item is generated by selecting randomly and with replacement values for $a(1)$ and $a(2)$ from the replacement set.

One advantage of the itemform approach is that each individual item need not be stored in its entirety; another, that itemforms are easily programmed on a computer. The primary disadvantage of this approach is that it is applicable only to those content areas where the subject matter is highly structured. Because of this, one expects and finds itemforms that have been used in elementary arithmetic, in elementary statistics, in spelling, and in science and mathematics (Shoemaker geeft literatuurverwijzingen hierbij).'

Opmerkingen. De itemforms approach zoals door Shoemaker beschreven levert weliswaar leuke itemconstructie regels op, maar lijdt overigens aan hetzelfde euvel waar alle traditionele item constructie gebruiken aan leiden: de wijze van formulering van de vragen steunt niet op een behoorlijke theorie (eer praktisch bruikbare theorie is ook voorlopig wel voldoende) die expliciet waarborgt dat cognitieve operaties die het doel van het onderwijs zijn, ook als zodanig in de vraagconstructie opgenomen worden: niet meer, en ook niet minder dan dat. Mijn verwachting is nu juist dat er regels gevonden kunnen worden die de door Shoemaker zo voorgestane isomorfie tussen onderwijsprogramma en toetsing (of tussen doelstellingen en toetsing) bewerkstelligen, en dat dergelijke regels de item forms kunnen genereren.

Transformational grammar approach.

(de benadering van Bormuth (1970), en in verbeterde versie Finn (1973) voor bibliografie zie Shoemaker).

"Although the transformational grammar approach does define unambiguously the item universe associated with the instructional program, this strategy is limited to those instructional programs whose content and substance can be stated explicitly in natural language. Additionally, in this approach, it is difficult to handle inferences or generalizations from content. Bormuth's approach has been criticized strongly by Diederich (1970) and to a lesser degree by Cronbach (1970). Cronbach, possibly realizing to a greater extent than Diederich what Bormuth was ultimately striving to achieve, commented that 'In the end, Bormuth's essay seems premature and disappointing, but if Daedalus comes, can the Wrights be far behind?'"

Een groot probleem voor de item pool constructie is het construeren van al die items: het laten doen door docenten zelf kost ontzettend veel tijd en energie waarbij ook nog eens veel items in tweede instantie niet geschikt blijken; het op commerciële basis laten construeren kost een hoop geld, wat niet beschikbaar kan zijn. Het is mijn voorlopige indruk dat al die kosten en moeiten voortvloeien wederom uit het ontbreken van een vraagformulerings-theorie: item schrijven is nog steeds een 'kunst', en 'kunstobjecten' zijn door hun zeldzaamheid kostbaar. Het fantastisch zwakke punt in de aanpak van Shoemaker is en blijft de wijze waarop vragen geformuleerd worden. Als je maar eenmaal een verzameling vragen hebt..... dan wordt alles makkelijker. Onze eerste zorg is nu juist het krijgen van voldoende greep op die vraagformuleringsproblematiek zelf, zodat iedere docent in korte tijd geoefend kan worden in het op onderwijskundig verantwoorde wijze in elkaar zetten van vragen, of dat nu meerkeuze vragen zijn, of vragen die een kort antwoord (aanvul- of open eindvragen) vereisen.

Fremer, J., & Anastasio, E. J. Computerassisted item writing. *Journal of Educational Measurement*, 1969, 69-74.

"The research reported here represents the first part of a project which aims to explore the itemwriting process, with a view toward developing computer techniques that would make possible the generation of verbal items. Objectives include the identification of words and sentences that are most relevant to the creation of verbal items, and the development of rules for the coding of words and sentences so that they can be manipulated by the computer.

There have been a number of recent indications that the time has arrived to employ the computer as an item writer. Osburn (1966), for example, has developed a computer procedure for writing statistics problems. He uses a series of 'question frames' (itemforms zou Shoemaker zeggen?) containing fixed and variable components. Each time a change is made in a variable component, a new item is created. Some resemblance can be seen between Osburn's approach and that used by Guttman (1965) in his facet analysis.⁰

De aanpak van Fremer & Anastasio lijkt op voorhand voor de meeste toetsingssituaties in het hoger onderwijs niet relevant. Volledigheidshalve neem ik de referentie op.

Wat wel van belang lijkt, is de verwijzing naar Guttman, wat oak doet vermoeden dat bij het maken van een item verzameling de generaliseerbaarheidstheorie van Cronbach, Gleser, Nanda, & Rajaratnam een belangrijk hulpmiddel bij het verduidelijken van de issues kan zijn. Mogelijk is dat zelfs ook al het geval waar het gaat om het zoeken naar regels voor het construeren van items, omdat die constructieregels betrekking hebben op domeinen van cognitieve activiteiten die in de doelstellingen van het onderwijs zijn opgenomen, en waarnaar men wil kunnen generaliseren op grond van toetsingsresultaten.

Wesman, A. G. Writing the test item. In Thorndike, R. L. (editor) Educational Measurement. Washington: American Council on Education, 1971

"Item writing is essentially creative -- it is an art." "Just as there

can be no set of formulas for producing a good story or a good painting, so there can be no set of rules that guarantees the production of good items."

Opmerkingen: met dit soort vooropgezette meningen wordt bij voorbaat het zoeken naar heuristieken voor constructie van items bijzonder onwaarschijnlijk gemaakt. Van de student uit gezien: wat heeft hij te maken met kunstmakers die wel eens even over zijn toekomst zullen beslissen? Als het schrijven van items een 'kunst' is, is het toch niet ver gezocht om het kunnen beantwoorden ervan ook als een 'kunst' te beschouwen? Wat moeten we met dit alles in het onderwijs?

"Each item, As it is being written, presents new problems and new opportunities." Geen ideale situatie, zou je zeggen. Gevolg daarvan kan bijna niet anders zijn dan dat ook de student bij ieder nieuw item met andere, onverwachte, ondoorzichtige trucs geconfronteerd wordt, een ongewenste situatie.

"Thus item writing requires an uncommon combination of special abilities and is mastered only through extensive and critically supervised practice." Onderwijs in de middeleeuwen zat ook ongeveer zo in elkaar.

"Item writing demands and tends to develop high standards of quality and a sense of pride in craftsmanship." En meer van dat soort snurkerige opmerkingen ontsieren de bijdrage van de heer Wesman.

Dat het schrijven van items vandaag de dag nog meer een kunst dan een kunde is, is wel ongeveer juist. In plaats van deze feitelijke toestand ideaal te verklaren, zou het beter geweest zijn op te merken dat juist hier een fantastische leemte in onze onderwijskundige kennis blijkt, een tekort dat betekent dat alle toetsing in het onderwijs een activiteit heeft met zijn grondslagen gebouwd op drijfzand. Daar helpen geen psychometrische trucs aan!

Het zoeken naar doelstelling-formulerings-technieken, waarmee

een aanzienlijk deel van de onderwijskunde vandaag de dag gepreoccupeerd is, kan gezien worden als een grootscheepse poging om de 'kunsten' van Wesman eindelijk eens te vervangen door item-schrijftechnieken met een rationele grondslag. Vandaar ook de niet geringe nadruk in de voorstellen voor een cognitieve taxonomie van onderwijsdoelstellingen (Bloom c.s., 1956) op illustratie met behulp van toets items.

De item schrijver moet grondig thuis zijn in het vakgebied waarover vragen geformuleerd worden. Wesman, en met hem vele anderen, doet er nog een schepje bovenop door te stellen dat de itemschrijver "should be aware of popular fallacies and misconceptions in the field. This is particularly necessary in the construction of multiplechoice items where a portion of the art of item writing is the ability to conceive of distracters which are incorrect by any standard and yet are attractive to those not knowing the correct answer." Opmerking: het kan nauwelijks de bedoeling van de gewone studietoets zijn om na te gaan of de leerlingstudent het onderwijs wel gevolgd heeft: gebruik daar liever een presentielijst o.i.d. voor. Desnoods de techniek van de 'onbenullige items', maar dat zijn nu juist items die iedere docent over zijn eigen vakmin hoog tempo kan schrijven (Hofstee, 1973).

En welk een opmerkelijke gedachte over de rol van de afleiders: niet alleen zouden ze zo moeten zijn dat de naïeve student (die niet aan het onderwijs heeft deelgenomen) met meer dan kans een fout alternatief aanstreept, óók de student die het onderwijs behoorlijk gevolgd heeft maar niet zeker is van het antwoord op een bepaalde vraag, zou dan verleid moeten worden om een onjuist alternatief aan te strepen! Deze idee is des te opmerkelijker omdat in de praktijk altijd weer blijkt dat partiële kennis niet leidt tot een lagere score (dan bij random invullen van de betreffende vragen het geval zou zijn), zoals de gedachtegang van Wesman vraagt, maar juist tot een hogere score (dan bij random invullen van de betreffende vragen). Dus, ook al zou de idee op zich aantrekkelijk zijn, in de praktijk is ze niet uitvoerbaar. En de idee is uiterst onaantrekkelijk: het is een krom soort ethiek studenten in beoordelingssituaties met trucs, strikvragen e.d. te lijf gaan tot summum van itemschrijfkunst verheft. Tenslotte: als er al zo

nodig meerkeuze items geschreven moeten worden. Niet voor de meeste onderwijssituaties waar docenten zelf hun toetsen maken niet de beste vraagvorm is, zie Wilbrink 5-3-76), dan moet de keuze van afleiders op onderwijskundige argumenten plaats vinden (zie Wilbrink 13-2-76).

"Every test begins with an idea in the mind of the item writer. The production and selection of ideas upon which test items may be based is one of the most difficult problems facing him. (...) If expressed formally, the idea for a test item would consist of a statement identifying some knowledge, understanding, ability or characteristic reaction of the examinee.(...) There is no automatic process for the production of item ideas. They must be invented or discovered, and in these processes chance thoughts and inspirations are very important." (onderstrepingen van mij, b.w.)

"One source that may stimulate the production of item ideas is provided by the written work of the students themselves. Their expressions of ideas on issues and problems may reveal points of difficulty that can be made the basis of discriminating test items."

Opmerkingen: de item schrijver is in de visie van Wesman al zo ver geprostitueerd door de psychometrie dat er al bij voorbaat vanuit gegaan wordt dat item in een studietoets moeten discrimineren. en dat punten waar studenten kennelijk moeite mee hebben gebruikt moeten worden ten dienste van die discriminatie en niet als aanwijzingen dat op die punten het onderwijs kennelijk niet goed geweest is en dat er dus niet op getoetst mag worden, maar het onderwijs op die punten eerst verbeterd moet worden!

We zien telkens weer geïllustreerd dat het doel van de toetsing volkomen buiten discussie blijft: aangenomen wordt dat toetsing ten doel heeft uit te maken welke leerlingen of studenten méér hebben opgestoken (of altijd al wisten wanneer de toets nog verder afzakt in de richting van test van intelligentie i.p.v. in het onderwijs opgedane kennis en vaardigheden) dan anderen. Juist in het onderwijs een irrelevante, en vaak zelfs kwaadaardige doelstelling voor beoordeling van prestaties. (kwaadaardig in de betekenis van het onderwijsschade berokkenend, de leerlingen vernederend, e.d.)

"Finally, in the selection of item ideas it is necessary to consider

their probable ability to discriminate between those who possess and those who lack a given understanding or ability." Wie denkt dat Wesman hier op het spoor van criterium gerefereerd meten zit, wordt snel ontnuchterd door de er onmiddellijk op aansluitende zin: "The chief purpose of most tests of educational achievement is to rank examinees as accurately as possible in order of their attainments." En je hoort Wesman er bij denken: de meeste toetsen zijn zo bedoeld, en zo horen toetsen ook bedoeld te zijn.

Tenslotte merkt Wesman naar aanleiding van mastery en diagnostische toetsen nog op dat voor dergelijke toetsen "psychometric standards with respect to desired difficulty and discrimination characteristics of items properly may be ignored". Niet meer dan lipservice aan enkele belangrijke functies van toetsen in het onderwijsproces. Merk op dat Wesman hier de constructie maakt dat 'psychometrische kwaliteit' iets is dat bestaat los van het doel waarvoor een toets gebruikt moet worden. Een aanfluiting dus. En hij doet het nog een keer, enkele regels verder: "As noted above, the educational decision should supersede the psychometric standard. It would be well, of course, for the educator who makes these decisions to recognize that he ordinarily may not ask that the test function as well for ranking purposes as it could if all items met satisfactory standards of psychometric quality."

Cronbach, L. J. Test validation. In Thorndike, R. L. (Editor) Educational Measurement. Washington D. C.: American Council on Education, 1971.

Voor de problematiek rond het formuleren van vragen kan het verhelderend zijn eens na te trekken wat diverse auteurs over de validiteitsproblematiek van toetsen te berde brengen: we kunnen immers een enkele vraag opvatten als een toets op zich.

"Professional constructors of achievement tests cross a content outline with a set of response process categories, expressing the latter in such terms as recall, reasoning, and application of principles. Such a specification has value in broadening the test. But, since task operations controlled by the tester are not to be

confused with processes used by the subject, it is evident that the usual content-by-process grid is not a universe specification. (voor de betekenis van de term 'universe specification' zie Cronbach's artikel, of ook Cronbach, Gleser, Nanda, & Rajaratnam 1972). An item qua item cannot be matched with a single behavioral process. Finding the answer calls for dozens of processes, from hearing the directions to complex integration of ideas. The shorthand description in terms of a single process is justified when one is certain that every person can and will carry out all the required processes save one. Even to speak of 'required processes', however, is misleading, since the task can perhaps be performed successfully by alternative processes (French, J. W. The relationship of problem-solving styles to the factor composition of tests. *Educational and Psychological Measurement*, 1965, 25, 928)"

"A proper response specification deals with the result a person is asked to produce, not the process(es) by which he succeeds or fails. 'Reads printed words aloud' is a description of an observable response; it says nothing about whether the reader is to look and say or to sound the word out. A person who insisted on separating these two response processes would have to devise a new task specification, perhaps requiring the reading of nonsense constructions that no subject has seen before. If a category of the form, say, 'ability to evaluate arguments' is to mean anything as a task specification, the designation must be fleshed out to describe something about the stimulus, the accompanying injunction to the subject, and the aspect of behavior to which the scorer is directed to attend."

Opmerking: met name waar het gaat om meerkeuze studietoets items wordt de door Cronbach gegeven regel veelal met voeten getreden: het van de student gevraagde gedrag wordt meestal onbeschreven gelaten, met als gevolg een ongecontroleerde, en daardoor vaak slechte, 'fit' van de vragen bij het onderwijs (of van de vragen met de doelstellingen). Natuurlijk is een triviale formulering als: de student moet uit meerdere alternatieven het juiste of het beste alternatief aanstrepen, geen vervanging voor een behoorlijke beschrijving van de gevraagde 'response'.

"For a multiple choice test, the specifications should go beyond a statement of topical categories to state the general characteristics of misleading alternatives (distractors). For some skill tests, it will even be possible to list halfadozen categories of errors that are to be systematically represented among the distractors. The content of an item can be altered radically by changing the distractors, while keeping the correct response the same. An arithmetic test may offer numerical distractors all of which are close to the correct answer, or it may offer a choice among widely scattered numbers. This changes what the test measures, as Messick and Kogan showed (Category width and quantitative aptitude. Perceptual and Motor Skills, 1965, 20, 493497)"

"The classification of objectives attempted in the Taxonomy of Educational Objectives (Bloom, 1956) does not adhere to the standard herein proposed. The authors stated that classification is primarily educational and logical, with psychological considerations subordinate. Yet the Taxonomy deals with 'The intended behavior of students the ways in which individuals are to act, think, or feel'."

"The headings of the Taxonomy may be taken as working hypotheses for the psychological interpretation of test behavior. What processes a test measures can be established only by construct validation; even then, the conclusion applies only to a certain population of persons."

"Content validity necessarily is limited by the adequacy of the universe specification, which is usually couched in imprecise everyday terms and can rarely mention every pertinent aspect of the task. R. B. Miller (1962, esp. pp. 193215) has provided a valuable discussion of task description, in the ideal case and as currently practiced. He noted that one must fall back upon a theory of task performance in deciding which cues or actions are relevant and therefore need to be specified. Hence, though psychological processes are not properly part of the task description, whatever one knows or suspects about them determines the way he specifies the universe."

"What, now, would constitute a rigorous validation of the

fit between the operational definition of the universe and the actual test operations? To stimulate thought, one can suggest an experimental validation through duplicate construction. The construction would involve judgment, but the validation would employ completely hard data. No study such as outlined here has ever been carried out, though Ebel (1962) has approximated it. While studies of this sort will surely provide disturbing results, they have much to teach regarding the preparation of instructions for item writers."

"The test constructor who takes content validity seriously will begin, not by jotting down items as they occur to him, or by soberly preparing one item regarding each page of the textbook, but by translating each of his educational objectives into task specifications. He usually will be most successful if he subdivides his gross objective rather finely. Within each subdivision he will need to go beyond listing topics that might be tested, to state the forms in which behavior relevant to each topic appears. He needs to clarify whether the student is to be presented with technical words, descriptions of situations in everyday words, pictures of situations, or concrete objects. Response specifications are equally important. All too often, the multiple-choice format has been taken for granted. If the constructor thinks about his aims with an open mind he will often decide that the task should call for a response constructed by the student—written open-ended responses or, if inhibitions are to be minimized, oral responses."

"The universe definition, once completed, goes a long way toward generating the items to be used or the observations to be made (onderstreping van mij, b.w.). According to the rationale for content validity, substituting other types of tasks for those the specifications call for cannot be justified by the argument that the substitutes correlate highly with the specified items. This evidence is pertinent enough when one is to measure individual differences; but substituting recognition items for recall items or verbal items for pictorial items probably alters the absolute score, and thus does not report level of mastery of the specified universe."

Stanley, J. C., & Hopkins, K. D. Educational and psychological measurement and evaluation. Englewood Cliffs, N.J.: Prentice Hall, 1972.

Een obligaat hoofdstuk over doelstellingen (Bloom, toestanden), echter zonder concrete uitwerking naar constructieregels voor toetsvragen, is het enige dat er in deze tekst te vinden is m.b.t. vraagconstructie. W&l uitgebreide aandacht aan allerlei formuleringstrucs, fouten die bij de formulering gemaakt kunnen, en dus vermeden moeten worden, e.d.

Ebel, R. L, Essentials of educational measurement. Englewood Cliffs, N. J PrenticeHall, 1972.

Weer een bekend tekstboek, waarin aan de inhoudelijke constructie van toetsvragen volstrekt onvoldoende aandacht gegeven wordt. Ebel doet twee suggesties: uitspraken in instructiemateriaal als uitgangspunt nemen (nogal nalef), en nieuwe toepassingen e.d. vragen (ook zo' globaal geformuleerd dat het niet interessant meer is):

"Good materials of instruction texts, references, lecture notes, and so forth should provide a convenient source from which to select key statements for conversion into test items." Opmerking: de last van inhoudelijke validering van je toetsvragen schuift Ebel hier lichthartig van zich af op de nek van een tekst schrijver die zich waarschijnlijk ook niet expliciet bezonnen heeft op zijn doelstellingen, en zeker niet op het nog weer geheel Andere probleem hoe je bepaalde doelstellingen in toetsvragen kunt 'vangen'. more "Laws, principles, and generalizations tend to be Important than specific incidents or details. However, the statement of such laws, principles, and generalizations may be memorized as relatively meaningless word sequences. This calls attention to one danger in basing multiplechoice test items on statements identical with those that have been used in instruction. The wording of the item may duplicate the statement so closely that verbal memory could provide as good a basis for correct response as would clear understanding. Good item writing avoids this danger by

rephrasing the idea or seeking novel illustrations or applications of it. The aim of good item writing is to yield evidence of understanding, of command of the idea as a tool in thinking, rather than evidence of recall of the conventional verbal expression of the idea." Hoe je een en ander moet bereiken, tja, dat zal in de ogen van Ebel ook wel een kwestie van 'kunst', 'intuïtie' men 'ervaring' zijn.

Thorndike, R. L., & Hagen, E. Measurement and evaluation in psychology and education. London: Wiley, 1969.

Niets over het inhoudelijke vraagconstructieprobleem, behalve natuurlijk het verplichte nummertje over formulering van doelstellingen.

Mehrens, W. A., & Lehmann, I. J. Measurement and evaluation in education and psychology. London: Holt, Rinehart, & Winston, 1973.

Bevat iets meer dan de drie al genoemde tekstboeken, namelijk een summiere verwijzing naar Hively, W., Patterson, H..L., & Page, S. H. A 'universedefined' system of arithmetic achievement tests. Journal of educational measurement, 1968, 2, 275-290. En Osburn, H. G. Item sampling for achievement testing. Educational and psychological Measurement, 1968, 28, 951-104.

Osburn, H. G. Item sampling for achievement testing. Educational and Psychological Measurement. 1968, 28, 951-104.

Een bijzonder aardig artikel. De idee lijkt als twee druppels water op wat ik voor de toetsconstructie moduul ontwikkeld heb (eerste twee voortgangsrapporten), zij het dat Osburn enerzijds de onderwijskundige kant volledig buiten beschouwing laat (of, zo je wilt, de cognitieve kant) met zijn onderscheidingen tussen verschillende item vormen, en anderzijds maakt Osburn een wat typische, hiërarchische, uitwerking van de 'itemform' idee. Hoe een en ander ook zij, dit werk, en ipso facto ook dat van Hively waarop

het zich laat inspireren, wordt pas relevant op het moment dat de vraagvorm vastgesteld is. Voor het bepalen van de vraagvorm zelf hebben we niets aan Osburn.

Merrill, M. D., & Boutwell, R. C. Instructional development: methodology and research. In Kerlinger, F. N. (editor) Review of Research in education, volume 1. Itasca, Ill.: Peacock, 1973.

Blz. 97 t/m 104 hierna gecopieerd (niet gescand, dat gaat teveel tijd kosten).

Klausmeier, H. J., Ghatula, E. S., & Frayer, D. A. Conceptual learning and development: a cognitive view. London: Academic Press, 1974.

In dit werk wordt een benadering van het vraagconstructie probleem gepresenteerd, die het meest vruchtbaar lijkt voor een praktische, heuristische, oplossing. De volgende, gecopiëerde, bladzijden 249-253 illustreren de zaak voldoende. De auteurs voegen er nog aan toe:

"The paradigm permits the generation of objectives and of related items other than the ones given. For example, items for objectives 4 and 5 may be written for all the attributes of quadrilaterals. Similarly, items for objective 11 could be written for other kinds of quadrilaterals. Objectives 10-13 could be eliminated if one were concerned only with the level of mastery of a particular concept and not with the uses of attained concepts.

It is noted that some concepts are not defined in terms of attributes. In this case, objectives 47 are not appropriate. For example, when a concept is defined in terms of synonyms or antonyms (beautiful pretty, beautiful ugly), the synonym or antonym may be substituted for the attribute."

Bloom, B. S. (Editor) Taxonomy of educational objectives. The

classification of educational goals. Handbook I: Cognitive domain. New York: McKay, 1956.

Zie ook de al gereleveerde kritieken van Shoemaker en van Cronbach.

De nadruk ligt op doelstellingen, waarbij gepoogd is het hele cognitieve gebied in een (hiërarchisch opgebouwd) classificatie schema te vangen. De relevantie voor ons onderwerp:

"The Handbook includes constructive suggestions for measuring each class of objectives and offers a number of examples of the different item types which have been used by examiners."

1.10 KNOWLEDGE OF SPECIFICS. The recall of specific and isolable bits of information. This refers primarily to what might be called the hard core of facts or information in each field of knowledge. We zijn vooral benieuwd wat er zoal met 'feiten' en 'informatie' bedoeld kan worden. De categorie wordt in de volgende onderdelen opgesplitst:

1.11 Knowledge of terminology. Knowledge of the referents for specific verbal and nonverbal symbols. "The learner must become cognizant of these terms and symbols and must learn the generally accepted definitions or meanings to be attached. " Dat betekent zo ongeveer dat de student de terminologie voor gebruik beschikbaar moet hebben (de termen kan produceren wanneer dat te pas komt bijvoorbeeld), en dat hij definities en/of beschrijvingen, eventueel in eigen bewoordingen, moet kunnen geven. Het is de vraag of doelstellingen die inderdaad niet verder gaan dan het 'kunnen geven van de definitie' wel in het onderwijs thuis horen; omdat onze probleemstelling echter niet het formuleren van onderwijsdoelen is, maar het vinden van toetsvormen voor onderwijsdoelen, zullen we dit soort kritieken i.h.a. proberen te onderdrukken.

Voor dit soort doelstelling lijken de volgende toetsvormen van belang: gegeven de naam van het begrip, de definitie, of een beschrijving kunnen produceren (of herkennen). Kennelijk

bedoelen Bloom c.s. toch ook dat gegeven een voorbeeld van het begrip, de student de naam kan geven; d.w.z. gegeven een voorbeeld of een nietvoorbeeld van een begrip kan de student het als zodanig ook thuis brengen (al dan niet gegeven de naam van het begrip, wat de moeilijkheidsgraad van de vraag kan beïnvloeden). Zo opgevat echter gaan we al aanzienlijk het beperkte wereldje van de 'kennis' zoals in dagelijks taalgebruik opgevat, te buiten: genoemd soort 'kennis' is immers van zeer groot belang binnen het onderwijs, en kan zelfs tot de allerhoogste 'bloomiaanse' cognitieve niveaus horen wanneer het gaat om begrippen die zeer 'complex zijn (denk aan medische diagnose bijvoorbeeld). Andere mogelijkheden: gegeven een beschrijving in nieuwe woorden, moet de student de naam van het bedoelde begrip kunnen geven. Etcetera.

Kennelijk vallen 'regels' niet onder deze categorie: regels bevatten meer 'bits' informatie.

1.12 Knowledge of specific facts.

Knowledge of dates, events, persons, places, sources of information, etc. Knowledge of specific facts refers to those facts which can be isolated as separate, discrete elements in contrast to those which can only be known in a larger context.

Deze categorie is een rommelpot van van alles wat. Er worden begrippen ingestopt waarvan de student eigenschappen moet kennen. Mijn indruk is dat de commissie Bloom vooral begrippen voor ogen hebben gestaan die uit slechts één element bestaan (de Maan, minister president Drees, de watersnoodramp). Er lijkt op voorhand weinig reden te zijn om begrippen met meerdere elementen anders te behandelen dan begrippen met slechts één element.

Dan schijnt in deze categorie te vallen alles wat encyclopedisch is; dus informatie waarvan lijsten aan te leggen zijn, opsommingen te maken zijn, etc., inclusief lijsten van vindplaatsen van andere lijsten (dat is: vindplaatsen kennen, bijv.: 'Knowledge of reliable

sources of' information for wise purchasing'.

Het lijkt dan ook dat deze categorie ofwel niets toevoegt aan andere onderscheiden categorieën, ofwel doelstellingen bevat die in het onderwijs aan de orde komen in de vorm van Handelingsonderdelen (en niet in formatieve of summatieve toetsing) of in de vorm van te raadplegen lijsten (die ook bij toetsing raadpleegbaar zouden kunnen of liever moeten zijn: open boek toestanden dus).

1.20 'KNOWLEDGE OF WAYS AND MEANS OF DEALING WITH SPECIFICS.

Knowledge of the ways of organizing, studying, judging, and criticizing ideas and phenomena. It does not involve actual use of the ways and means so much as it does a knowledge of their existence and possible use.

1.21 Knowledge of conventions. Knowledge of characteristic ways of treating and presenting ideas and phenomena.

De auteurs geven zelf al aan dat categorie 1.20 moeilijk te onderscheiden is van 1.10, en da.ar valt niets meer aan toe te voegen.

1.22 Knowledge of trends and sequences. Knowledge of the processes, directions, and movements of phenomena with respect to time.

idem dito van hetzelfde.

1.23 Knowledge of classifications and categories.

1.24 Knowledge of criteria.

Het probleem met deze categorieën wordt aardig aangeduid in de volgende passage: "They (de criteria) are likely to appear complex and abstract to students and to acquire meaning only

as they are related to concrete situations and problems" welk laatste gebeurt in cat. 6.00 'evaluatie'. Mijn vraag is dan telkens weer: waarom kennelijk voor het onderwijs irrelevante doelen toch een plaatsje inruimen in een 'classificatie systeem'. Waarom niet liever bijv. een doelstelling die zou vallen onder 1.2 beschouwd als onderdeel van doelstellingen thuishorend onder 6.00 evaluatie?

1.25 Knowledge of methodology.

Hetzelfde probleem wederom.

1.30 KNOWLEDGE OF THE UNIVERSALS AND ABSTRACTIONS IN A FIELD.

Knowledge of the major ideas, schemes and patterns by which phenomena and ideas are organized. These are the large structures theories and generalizations which dominate a field or which are quite generally used in studying phenomena or solving problems. These are at the highest levels of abstraction and complexity.

These concepts bring together a large number of specific facts and events, describe the processes and interrelations among these specifics, and thus enable the worker to organize the whole in a parsimonious form.

These tend to be very broad ideas and plans which are rather difficult for students to comprehend. Quite frequently they are so difficult because the student is not thoroughly acquainted with the phenomena the universals are intended to summarize and organize. If the student does get to know them, however, he has a means of relating and organizing a great deal of subject matter and as a result should have more insight into the field as well as greater retentiveness for it.

Well, wat hebben we hier anders dan de regels van Gagné, al dan niet door andere regels nog weer aan elkaar gerelateerd in georganiseerde, voor mijn part hiërarchische, structuren? Misschien is een pertinente vraag hier of het nodig is in de doelstellingen t.b.v. de toetsing nog verder te, gaan dan uitsluitend

de begrippen, regels, eventueel 'hyper regels', waaruit de leerstof is opgebouwd.

1.31 Knowledge of principles and generalizations. Knowledge of particular abstractions which summarize observations or phenomena.

Bedoeld worden hier het soort regels die (onder categorie 3.00) toepasbaar zijn.

T32 Knowledge of theories and structures.

2.00 COMPREHENSION That is, when students are confronted with a communication,

they are expected to know what is being communicated and be able

to make some use of the material or ideas contained in it.

Here we are using the term 'comprehension' to include those objectives, behaviors, or responses which represent an understanding of the literal message contained in a communication.

Een soort afval categorie voor doelstellingen die op de een of andere manier te simpel zijn voor hogere doelstellingscategorieën. Ik kan er dan ook weinig zinnigs mee aanvangen, en het lijkt me in haar geheel een behoorlijk overbodige categorie. Het is een soort slechte erfenis voortvloeiend uit de miskleun van louter verbale napraterij als een doelstellingscategorie op zich (1.00) te beschouwen.

2.10 Translation. Ofwel in eigen woorden weer kunnen geven, of in andere symboliek kunnen vertalen.,

2.20 Interpretation. Datgene nl. wat gebeurt wanneer meerdere vertalingen tegelijk moeten worden gewaakt (wanneer dus van meerdere begrippen tegelijk sprake is, bijv.)

2.30 Extrapolation. De scheiding met hogere categorieën

wordt hier heel erg schimmig.

Dat categorie 2.00 ook erg moeilijk te onderscheiden is van de kennis categorie wordt wel duidelijk gemaakt door de volgende uitspraak:

"Student ability to translate formal definitions or statements of principles may also be evaluated by exercises requiring him to recall or recognise correct illustrations or examples." "Waarbij zelfs kennelijk sprake is van voorbeelden die identiek zijn aan reeds in het onderwijs behandelde.

"For example, terms may be listed and their definitions requested, or the definitions may be given and the student merely asked to recall or recognize the appropriate terms."

3.00 APPLICATION.

The whole cognitive domain of the taxonomy is arranged in a hierarchy, that is, each classification within it demands the skills and abilities which are lower in the classification order. The application category follows this rule in that to apply something requires 'Comprehension' of the method, theory, principle or abstraction applied. Teachers frequently say, 'If a student really comprehends something, then he can apply it.'

Waarschijnlijk hebben die 'teachers' gelijk, en is het inderdaad niet zinvol te spreken van 'begrijpen' zonder het 'kunnen toepassen'. Waarbij onder toepassen dan vallen: het classificeren van voorbeelden van een begrip en nietvoorbeelden van een begrip als al dan niet behorend tot het begrip; het uitvoeren van procedures (regels); het aan kunnen geven op welke instanties een bepaalde regel toepasselijk is of niet, etcetera. Dat alles waarschijnlijk in opgaven waarin niet van de student verlangd wordt om ook zelf uit te zoeken om welke regels of begrippen het gaat (dat zou in hogere categorieën aan de orde moeten komen ?) Dit laatste punt is niet helemaal duidelijk, waar Bloom c.s. zeggen:

"A demonstration of 'Comprehension' shows that the student

can use the abstraction when its use is specified. A demonstration of 'Application' shows that he will use it correctly, given an appropriate situation in which no mode of solution is specified."

Het laatste mag dan zo' vertaald worden: bij toepassing van regels. is van toepassing in de zin van categorie 3.00 sprake wanneer de student, geconfronteerd met een voorbeeld van een of het begrip waarop de regel van toepassing is, de juiste regel juist toepast.

4.00 ANALYSIS.

Tekstanalyse is een activiteit waarbij gebruik gemaakt wordt van een aantal begrippen en regels, als geïllustreerd in het volgende:

"(teachers) wish, for example, to develop in students the ability to distinguish fact from hypothesis in a communication, to identify conclusions and supporting statements, to distinguish relevant from extraneous material, to note how one idea relates to another, to see what unstated assumptions are involved in what is said, to distinguish dominant from subordinate ideas or themes in poetry or music, to find evidence of the author's techniques and purposes, etc., etc."

De categorie 'Analyse' is eigenlijk slechts een categorie doelstellingen met betrekking tot een heel bepaalde onderwijsinhoud (vak). Het lijkt een heilloze weg om. voor afzonderlijke leerstofgebieden ook afzonderlijke doelstelling categorieën te gaan onderscheiden. Doelstellingen voor categorieën tekst analyse zijn niet essentieel verschillend van doelstellingen voor bijv. biologie, of wiskunde. [sic]

5.00 SYNTHESIS.

In synthesis the student must draw upon elements from many sources and put these together into a structure or pattern not clearly there before. His efforts should yield a product.

Een onderscheid kan dan nog gemaakt worden tussen het ontwerpen en het ook werkelijk uitvoeren; de uitvoering van al gemaakte plannen is op zich niet meer iets dan onder de categorie 'synthese' valt. Voor de toetsing is eigenlijk alleen de plannenmakerij, het construeren van een ontwerp, schema, aanpak, etc., van belang (althans, in die gevallen waar inderdaad van de concrete uitvoering afgezien kan worden).

Die 'toetsing' is wel problematisch: meestal immers zal het gaan om produkten die vrij veel tijd vergen, m.a.w. in summatieve toetsing zal het niet eenvoudig zijn om synthese doelstellingen te 'vangen'. Bovendien is het ook niet nodig, omdat deze categorie doelstellingen bij uitstek de Handelingen (De Groot 1972) zijn die tijdens het onderwijs aan de orde zijn, waar de student mee bezig is, of mee bezig geweest moet zijn, onder goede begeleiding uiteraard.

Er zijn een aantal doelstellingen die zich wat moeilijk laten onderscheiden van de categorie 'synthese'. Het maken van berekeningen, geven van bewijzen etc. zijn veelal activiteiten die vallen onder 'toepassing', zij het dat we dan de categorie 'toepassing' wat ruimer moeten nemen dan Bloom c.s. doen. Bij de meeste berekeningen, constructies, bewijsvoeringen e.d. gaat het immers om het toepassen van meerdere begrippen en regels, dus ook om het herkennen van situaties als geschikt om bepaalde regels in combinatie op toe te passen. Het soort combinatie van regels in deze toepassingen zal in het onderwijs doorgaans aan de orde zijn geweest: is dat niet het geval, dan zijn de problemen zo moeilijk dat ze niet in aanmerking komen voor opname in een summatieve toets vanwege het grote tijdsbeslag, of omdat ze op dat moment meer de intelligentie van de student dan zijn vakkennis testen. Het vinden van nieuwe soorten bewijsvoering of berekeningen of constructie etc. valt onder de categorie 'synthese', en, zoals gezegd gaat het hier om doelstellingen die De Groot zou aanduiden als Handelingsdoelen.

(Evenals bij categorie 4.00 zijn hier de onderverdelingen in deelcategorieën niet evident zinvol, en worden overgeslagen).

Zoals Bloom c.s. ook al opmerken (pp. 174) is een reden om

synthese doelstellingen niet te toetsen d.m.v. summatieve toetsen dat de producten moeilijk beoordeelbaar zijn: niet alleen moet er noodzakelijkerwijs nogal subjectief beoordeeld worden, maar daar komt bij dat het juist de bedoeling is dat er producten gemaakt worden die dermate 'vrij' zijn dat van tevoren niet aangegeven kan worden aan welke 'kwaliteiten' het antwoord 'scoorbaar' is (behalve aan de randvoorwaarden waarbinnen het product gemaakt moet worden, maar overschrijding van randvoorwaarden is een triviale kwestie in de totale 'beoordeling' van het product). Bloom c.s. leggen sterk de nadruk op het 'creatieve' van de synthese doelstellingen, en zodra het om werkelijk creatieve producten gaat is iedere beoordeling 'shaky'.

Om meer of minder dezelfde redenen vallen doelstellingen behorend tot de laatste categorie

6.00 EVALUATIE

ook buiten het bereik van summatieve toetsing. Een bespreking kan dan ook achterwege blijven.

Tot hier dan het werk van Bloom c.s. uit 1956.

Bloom⁷ B. S., Hastings, J. T., & Madaus, G. F. (editors) Handbook on formative and summative evaluation of learning. London: McGrawHill, 1971(i.h.b. section three: Evaluation techniques for cognitive and affective objectives.)

Over 'Kennis':

"Although many of the specifics which we learn to recall or recognize during our formal instruction are forgotten within a few months or years, knowledge of them at the time of learning is extremely important for the development of ideas which do stay with us for interpretive and associational uses. Therefore, during the instruction period it is important for the teacher to test such knowledge as one facet of the evaluation of student learning. In many cases assessment of recall of specific facts is more a function

of formative evaluation than of summative. We need to have evidence of whether the student can recall certain terms, facts, or methods in order to make inferences about difficulties or next treatments."

Bloom c.s. verhelderen hiermee veel van wat in de Taxonomie in een stortvloed van gepsychologiseer niet boven water kon komen. De les voor het vraagconstructie probleem is kennelijk dat er nogal wat vraagsoorten zijn die eigenlijk in een afsluitende toets misplaatst zijn (al te specifieke vragen over letterlijke teksten (altijd al uit den boze), maar ook definities van termen, beschrijving van begrippen of regels, en in sommige gevallen ook het thuisbrengen van voorbeelden van een begrip als behorend tot dat begrip, etc.).

Over 'Toepassing':

Benadrukt wordt dat toepassingen gevraagd worden voor nieuwe problemen en situaties. Bloom c.s. steken dan een heel verhaal af over wat 'nieuw' is, waarmee de zaakodeloos gecompliceerd gemaakt wordt:

"it is not a new problem or situation if it is exactly like others solved in class except that new quantities or symbols are used (as in mathematics or physics). It is not a new problem or situation if it is the same as one solved in class with only slight changes altering the original form.

It is a new problem if or situation if the student has not been given instruction or help on a given problem and he must do some of the following with the statement of the problem before he can solve it:

1. The statement of the problem must be modified in some way before it can be attacked.

2. The statement of the problem must be put in the form of some paradigm or model before the student can bring the principles or generalizations previously learned to bear on it. This is especially true of mathematics and science problems where the

student has previously been given problem types and models and some practice in using them. It is more difficult to find such forms and structures in the social sciences and humanities.

3. The statement of the problem requires the student to search through his memory of principles and generalizations to find those which are relevant to it.(....)

"The posing of new problems and situations is a difficult art in evaluation." Een uitspraak waarmee ik het grondig oneens wil zijn. De taak van de docent kan aanzienlijk verlicht worden wanneer hem een aantal heuristieken ter beschikking staan voor het genereren van toepassingsvragen, heuristieken die Bloom c.s. niet aanreiken, maar die zeer wel te ontwikkelen zijn op basis van bijv. het werk van Klausmeier c.s., of Merrill & Boutwell.

PRINCIPLES. By 'principle' we mean a statement about a process or relation which describes a fundamental truth, or a law accepted and used by the scholars in a field. (...) These principles are statements which hold under a great range of conditions. They represent relatively precise inferences from a large body of observations and experiments, or are deductions from a body of theory or accepted assumptions.

GENERALIZATIONS. By 'generalizations' we mean general statements or inferences which summarize a body of information or other particulars and which can be applied in new situations. Some examples are frustration increases anxiety; the educational aspirations of students are related to their parents' socioeconomic status.

A generalization holds under limited conditions and the special problem it poses to the student is to recognize its tentativeness, the particular circumstances under which it may hold true, and its value in quickly ordering new phenomena. One hopes that the student will be able to understand the basis for the generalization, the underlying phenomena at work, and the restricted truth which the generalization summarizes.

Bloom c.s. maken dan nog een kleine onderverdeling:

Principles and generalizations are developed in a number of different ways. Some are developed by definition and convention. Workers in the field agree on a rule, definition or relationship.

the circumference of a circle equals $\text{force} = \text{mass} \times \text{acceleration}$
friction exists between two bodies in contact with each other.

Such abstractions are conventions which are very effective in classifying and organizing a great many specific phenomena in their respective fields. There is an arbitrary quality about these generalizations since they are true only by definition. What the student must learn about them is what they include and what they do not include. He must also learn what specific phenomena, concrete objects, and events illustrate the rule and which ones do not. He may also be expected to know various relations among the abstractions whether the relations are chronological, superordinate, concomitant variation, or cause and effect. He may also be expected to know whether the abstractions are parts of a theory, taxonomy, or classification scheme.

Since these abstractions are true only by definition, it is not possible for the student to discover them on his own.

Another set of abstractions are the result of empirical research and scholarship. (...) These generalizations are true within certain limits and they can be used to organize many specific events, actions, facts, and so forth. Abstractions of this type may be 'discovered' afresh by each student if the phenomena and the necessary measuring and observational techniques are available.

A third type is a more limited set of generalizations derived from observation. The following are illustrations of this type:

smoking cigarettes may cause cancer
too rigid toilet training may produce personality disorders
high interest rates reduce capital investment.

These generalizations may be statements of trends, tendencies, and regularities which are far from universal laws and principles; that is, they may occur frequently but there are many exceptions.

They may be examples of an effect which may also be produced by quite different causes, or illustrations of phenomena which have multiple causation (a.*.) Such generalizations have so many exceptions that the student must learn to be cautious in using them. (..4.) Again, he must be given opportunities to relate the abstract statement to specific and concrete examples.

A final set represent generalizations which are deductions from a larger theory, model, or point of view. Such generalizations may rarely describe existing phenomena but are statements of what would occur under special circumstances or ideal conditions. Examples are:

If children are given sufficient love and attention, they 'will not develop anxiety ()

Population tends to multiply faster than its means of subsistence.

Such generalizations represent a particular (and somewhat restricted) view of the world, and they are true if a particular set of assumptions are accepted. In learning such generalizations the student must clearly recognize the point of view they represent, and he should be helped to understand that they represent one alternative when others may also exist.

THE ABILITY TO APPLY.

The 'ability to apply' implies that with appropriate training, practice, and other kinds of help the student becomes able to apply principles and generalizations to new problems and situations. That is, he can appropriately use principles and generalizations in solving problems that are new to him. The implication is that there is a generalized ability which grows out of the student's learning experiences such that, when faced with new problems and situations, he can make use of the ability to apply principles and generalizations.

Hot laatste is een stukje behoorlijke grensoverschrijding van de"

heren Bloom en anderen: er wordt heel makkelijk overgeschakeld van het kunnen toepassen van bepaalde, in het onderwijs aan de orde geweest zijnde, regels op een soort problemen en situaties dat ook in het onderwijs aan de orde is geweest, naar het vermogen om regels (in het algemeen) toe te passen op nieuwe problemen en situaties (van een soort die niet noodzakelijk in het onderwijs aan de orde geweest hoeft te zijn). Kortom, het bespaart de docent die het laatste O'Ok wil een hoop ellende wanneer hij, in plaats van met veel moeite een studietoets in elkaar te zetten, gewoon een algemene intelligentie test gebruikt.

Voor ons eigen gebruik zullen we in het volgende telkens veronderstellen dat waar gesproken wordt over regels (principles and generalizations) en nieuwe problemen en situaties, het gaat om regels en soorten problemen en situaties die in het onderwijs aan de orde zijn geweest.

In terms of student behaviors, the ability to apply might include some of the following:

the student can determine which principles or generalizations are appropriate or relevant in dealing with a new problem situation. (A)

The student can restate a problem so as to determine which principles or generalizations are necessary for its solution. (B)

The student can specify the limits within which a particular principle or generalization is true or relevant (C).

The student can recognize the exceptions to a particular generalization and the reasons for them. (D)

The student can explain new phenomena in terms of known principles or generalizations (E)

The student can predict what will happen in a new situation by the use of appropriate principles or generalizations (F)

The student can determine or justify a particular course of action or decision in a new situation by the use of appropriate principles and generalizations.(G)

The student can state the reasoning he employs to support the use of one or more principles or generalizations in a given problem situation (H).

Niet alleen. het onderscheid in verschillende soorten regels, maar ook het boven gegeven rijtje kan een belangrijke aanvulling zijn op het voorlopige schema zoals door mij in de eerste twee voortgangsrapporten moduul toetsconstructie uitgewerkt. Het hoofdstuk 'Toepassing' in dit handboek is in ieder geval een belangrijke stap vooruit vergeleken met het desbetreffende hoofdstuk in de oorspronkelijke Taxonomie.

Bloom c.s. geven uitgebreide voorbeelden bij bovenstaande lijst van bijbehorende toetsitems; die voorbeelden vragen nog om nadere bestudering (blz. 166-177) Vooral van belang bij de door Bloom c.s. gegeven voorbeelden is dat zij van een meerkeuze type zijn dat misschien minder of geheel niet geplaagd wordt door een aantal bekende nadelen van het traditionele type. Ook op dit aspect moeten de voorbeelden van items nog bestudeerd worden. Misschien vergis ik me, en bestaat ieder item gewoon uit een kleine collectie tweekeuze items o.i.d.

ANALYSE. Analyse blijft ook hier een activiteit die mijns inziens beter beschouwd kan worden als specifiek voor een bepaald vak, dan als algemeen te hanteren soort doelstelling. Dat vak, tekst analyse en interpretatie, is op zich natuurlijk wel belangrijk genoeg om misschien in een afzonderlijk hoofdstuk in het moduul toetsconstructie aandacht te besteden aan mogelijk specifieke vraag constructie problemen die hierbij opdoemen. In dat geval zal het betreffende hoofdstuk in Bloom c.s., met de daarin gegeven voorbeelden van items, zeer goed van pas kunnen komen.

SYNTHESE en EVALUATIE komen nauwelijks of niet voor summatieve toetsing in aanmerking:

Some of the requirements for developing synthesis problems and tests may be summarized briefly.

- The problem, task, or situation involving synthesis should be new or in some way different from those used in the instruction. The student may set the task or problem for himself or at least have considerable freedom in redefining it.

- The student may attack the problem with a variety of references or other materials available to him as he needs them. Thus, synthesis problems may be open book examinations, in which the student may use his notes, references, the library, and other resources as he finds them appropriate. Ideally synthesis problems should be as close as possible to the scholar (or artist, engineer, and so forth) attacking a problem he is interested in. The time allowed, conditions of work, and other stipulations should be as far from the typical, controlled examination situation as possible.

The adequacy of the final product may be judged in terms of the effect it has on the reader, observer, or audience; the adequacy with which it has accomplished the task; or evidence on the adequacy of the process by which it was developed.

Desalniettemin, ook al zijn doelstellingen die onder Synthese of Evaluatie vallen moeilijk of niet in een eindtoets te evalueren, toch zal de docent in de desbetreffende onderwijsonderdelen graag informatie verzamelen over de mate waarin het lukt deze doelstellingen, meestal in Handelingsonderdelen, gerealiseerd te krijgen. Het ziet er dan ook naar uit dat de beoordelingsproblematiek te behandelen in het moduul toetsconstructie een bijlage moet krijgen waarin beoordelingsproblemen rond synthese en evaluatie doelstellingen aan de orde gesteld worden. Te zijner tijd zal ik dus nog terug moeten komen op de behandeling van Bloom c.s. van deze problematiek.

afgesloten: 18 maart 1976.

projecten/76Vraagconstructie_8-3-76.pdf