

Noten

1. Dit onderzoek wordt uitgevoerd aan de Stichting Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam met subsidie van SVO.

Literatuur

- Siegel, S.
1956 *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

Analytische beoordeling van de samenvattingsofdracht C.S.E. Nederlands voor het VWO

Ton Hendrix en Piet Sanders

1. Inleiding

“Geef van de onderstaande tekst, die ongeveer 2200 woorden telt, een samenvatting in maximaal 500 woorden”.

Zo luidt de opdracht boven een betogende tekst die VWO-leerlingen in 2^{1/2} uur tijd voor hun C.S.E. Nederlands moeten samenvatten. Het is een eenvoudig geformuleerde opdracht voor een complexe taakstelling, zowel voor de examenkandidaat als voor de beoordelaars. In de 35 jaar dat de samenvattingsofdracht deel uit maakt van het C.S.E. voor het VWO (vroeger Gymnasium en HBS-b) is er al het nodige gezegd over de pedagogische waarde, over vakinhoudelijke, vakdidactische en – de laatste jaren – de toetstechnische problemen en onduidelijkheden van dit examenonderdeel.

In deze bijdrage zal verslag worden gedaan van een op het Cito verricht onderzoek naar de mogelijkheden de beoordelaarsovereenstemming bij de samenvattingsofdracht te verbeteren door middel van een daartoe ontwikkeld analytisch beoordelingsschema. Eerst komt de ontwikkeling van het tekstafhankelijk, analytisch beoordelingsschema dat bij het onderzoek gebruikt is aan de orde, daarna volgt een korte weergave van de opzet en de resultaten van het onderzoek.

2. De huidige situatie

Allereerst geef ik een overzicht van de huidige stand van zaken bij de beoordeling van de samenvatting op het VWO-eindexamen. De leerlingen krijgen een betogende tekst met de genoemde opdracht om daarvan een samenvatting te maken van maximaal 500 woorden. Ofschoon in de opdracht niet is aangegeven welk type samenvatting verlangd wordt, is zo langzamerhand wel duidelijk geworden uit het correctievoorschrift aan docenten, dat het gaat om wat Van Eemeren et al. (1975) een ‘informatieve samenvatting’ noemen: een complete tekst die aanzienlijk korter is dan de oorspronkelijke tekst en die de hoofdzaken en de hoofdlijnen daarvan weergeeft.

De docent-beoordelaar wordt geacht zijn oordeel over de samenvatting te baseren op de door de Centrale Examencommissie verstrekte *Aanwijzingen voor de beoordelaar*. De aanwijzingen komen op het volgende neer:

- De samenvatting moet uit maximaal 500 woorden bestaan. Afhankelijk van de overschrijding van dit maximum wordt één cijferpunt in mindering gebracht of wordt de samenvatting als onvoldoende beoordeeld.
- De samenvatting moet een duidelijk beeld geven van de gedachtengang en dus representatief zijn voor de tekst. De kandidaat moet tonen hoofd- en bijzaken te kunnen onderscheiden.
- De samengevatte tekst moet door iemand die de oorspronkelijke tekst niet kent, gevolgd kunnen worden.

Deze tekstonafhankelijke beoordelingscriteria gaan vergezeld van een tekststructuurschema dat door de examencommissie vrijblijvend aan de *Aanwijzingen voor de beoordeling* is toegevoegd. Bij dit tekststructuurschema wil ik even blijven stil staan, omdat dit als voorloper van ons analytisch beoordelingsschema beschouwd kan worden. Het is gebaseerd op de door Drop & De Vries (1980) ontwikkelde thematische analyse van informatieve teksten en op de opvattingen van Teun van Dijk over de macro-structuren van teksten (vgl. Van Dijk 1978). De bepaling van hoofd- en bijzaken geschiedt volgens de zogenaamde 'bottom-up-analysis' (Drop 1983). De leden van de examencommissie stellen in onderling overleg de hoofd- en subuitspraken van de tekst vast, alsmede de relaties daartussen. Het op deze wijze tot stand gekomen tekststructuurschema is aldus een weerspiegeling van de macro-structuur van de tekst.

Het is op dit moment niet duidelijk hoe dit tekststructuurschema door docenten gebruikt wordt. Sommige docenten hebben in de loop der jaren een eigen min of meer analytische beoordelingsmethode ontwikkeld en geven het tekststructuurschema daarbinnen een plaats, anderen kijken met een schuine blik naar het schema en bepalen hun cijfer volgens de globale beoordelingsmethode. Een voorbeeld van de eerste werkwijze is terug te vinden in het examennummer van *Levende talen*, waar C. Rooker (1984) uitvoerig verslag doet van het door hem ontwikkelde en gehanteerde beoordelingschema. Wel duidelijk is, dat de manier waarop het eindcijfer voor de kandidaten tot stand komt erg varieert en de nodige meningsverschillen en conflicten tussen eerste en tweede corrector tot gevolg heeft.

Om een meer uniforme beoordeling van de samenvatting mogelijk te maken, is onderzocht of de ontwikkeling van een analytisch beoordelingschema daartoe mogelijkheden biedt.

3. Criteria waaraan een samenvatting moet voldoen

Voor de ontwikkeling van zo'n analytisch beoordelingsschema is het noodzakelijk te weten welke criteria aan een samenvatting gesteld worden. Zowel uit publicaties over de samenvatting (Van Eemeren et al. 1975) als uit opiniepeilingen onder docenten (o.a. Kreeft & Luijten 1980) blijkt dat een samenvatting moet voldoen aan de volgende criteria, in volgorde van belangrijkheid:

1. de samenvatting dient een verkorte weergave te zijn van de hoofdzaken en hoofdlijnen van de oorspronkelijke tekst;

2. de samenvatting dient geen nieuwe elementen te bevatten, bijvoorbeeld in de vorm van commentaar op de oorspronkelijke tekst;
3. de samenvatting dient correct geformuleerd te zijn (dat wil zeggen 'in goed lopende zinnen');
4. de voorgeschreven lengte (meestal een kwart van de oorspronkelijke tekst) mag niet overschreden worden;
5. het ordeningsprincipe van de samenvatting dient ontleend te zijn aan de structuur van de oorspronkelijke tekst;
6. een 'eigen formulering' is gewenst, al is het toegestaan woorden en/of zinsneden letterlijk uit de oorspronkelijke tekst over te nemen;
7. de presentatie van de samenvatting (titelvermelding, auteursnaam, eventueel bronvermelding) moet direct inzichtelijk zijn voor de niet in de oorspronkelijke tekst ingewijde lezer.

Wanneer criterium 4 (tekstlengte) buiten beschouwing gelaten wordt, kunnen in deze criteria drie categorieën onderscheiden worden:

- a. Criteria die betrekking hebben op de inhoud van de samenvatting (criterium 1 en 2);
- b. Criteria die betrekking hebben op de formulering van de samenvatting (criterium 3 en 6);
- c. Criteria die betrekking hebben op de organisatie van de samenvatting (criterium 5 en 7).

4. Vooronderzoek met een uitgebreid schema

Op grond van bovenstaande indeling is er in eerste instantie een beoordelingsmodel opgesteld met een I-, F- en O-categorie (zie bijlage I). Om zo dicht mogelijk bij de huidige examenpraktijk te blijven, is er gekozen voor een scoreschaal van 100 punten minus 10 bonuspunten. Negentig scorepunten werden als volgt verdeeld over de drie categorieën: I - 60 punten; F - 25 en O - 15. (Deze verhouding was mede gebaseerd op gegevens uit een enquête onder docenten over wat zij belangrijk vonden in een goede samenvatting.)

Voor de beoordeling van de I-categorie werd gebruik gemaakt van het door de examencommissie opgestelde tekststructuurschema in nagenoeg ongewijzigde vorm. Door middel van weging van onderdelen werden de 60 scorepunten over de informatie-elementen verdeeld. De hoofduitspraken kregen maximaal 5 scorepunten, de subuitspraken 1 of 2. Het geven van deelscores (in hele punten) was mogelijk. De F- en O-categorie werden beoordeeld met behulp van een aantal beoordelings schalen. Met behulp van dit schema hebben 10 VWO-docenten 25 examenwerken beoordeeld. Bij analyse van de resultaten bleek het volgende:

- De I-categorie bleek de meest betrouwbare te zijn.
- De F-categorie interfereerde met de twee overige categorieën.
- De op de I-categorie uitgebrachte scores correleerden hoog met scores op de drie categorieën gezamenlijk.

Op grond van deze gegevens en de commentaren van de docent-beoordelaars werd besloten een meer vereenvoudigd beoordelingsschema (zie bijlage II) te ontwikkelen, waarbij vooral de nadruk kwam te liggen op de inhoudelijke weergave van de samen te vatten tekst.

5. Onderzoek met een vereenvoudigd schema

Uitgangspunt voor het nieuwe beoordelingsschema bleef het reeds eerder genoemde tekststructuurschema. De lijnstructuur in het schema, die de expliciete aanwezigere relatie weergeeft, werd losgelaten door de funktiewoorden (oorzaak, gevolg, tegenstelling, etc.) te vervangen door verbindingsworden (omdat, zodat, maar, etc.). Bovendien werd de schema-taal (normaliseren, substantiveren) weer uitgeschreven tot lopende zinnen. Om aan de formuleringcomponent van de samenvattingsoopdracht tegemoet te komen, werd een tussenweg bewandeld tussen een volledig uitgeschreven model-samenvatting en een strak tekststructuurschema. De relaties binnen de tekst en de formulering van de samenvatting kregen binnen het nieuwe schema weer een wat duidelijker plaats om de beoordelaar de gelegenheid te geven na te gaan of de samenvatting van de kandidaat de belangrijkste informatie-elementen en relaties daartussen bevatte. Dit analytisch beoordelingsschema is te beschouwen als een lijst van informatie-elementen waaraan door middel van weging van onderdelen een bepaalde score wordt toegekend. Onder een informatie-element wordt verstaan: een afgeronde bewering, con-statering of beschrijving die wordt uitgedrukt in één of meer zinnen. De in het schema opgenomen informatie-elementen representeren de hoofdlijnen van het door de auteur gegeven betoog in de volgorde waarin hij het de lezer aanbiedt.

- Kenmerkend voor het nieuwe schema was het volgende:
 - Het scoren van het examenwerk was vooral afgestemd op de inhouds-component.
 - Er bestond ook de mogelijkheid van het geven van aftrekpunten voor volgens het beoordelingsschema overbodige informatie-elementen. Dit om te voorkomen dat te uitgebreide samenvattingen - waarin overigens alle noodzakelijke informatie-elementen ook vermeld waren - een te hoge score zouden krijgen.
 - De hoofduitspraak en de bijbehorende subuitspraken (Drop & De Vries 1980) vormden samen een beoordelingseenheid. De voor het onderzoek gebruikte tekst viel zo uiteen in zes beoordelingseenheden.
 - Aan elk informatie-element was een maximum-score toegekend. Dit liet de beoordelaar de ruimte om geen punten toe te kennen bij afwezigheid of foutieve weergave van het informatie-element en slechts een gedeelte van het totaal te behalen punten bij onvolkomenheden en/of onvolledige weergave van zowel de informatie-elementen als de relaties tussen de tekstelementen.

- Bij onzorgvuldig en onjuist woordgebruik dienden bij het scoren van de informatie-elementen alleen dan punten in mindering te worden gebracht, wanneer de samengevatte tekst niet begrijpelijk was voor iemand die de uitgangstekst niet gelezen had.

Tenslotte gold dat het beoordelingsschema als een bindend correctievoorschrift moest worden gehanteerd.

6. Uitvoering van het onderzoek

Voor het onderzoek met het vereenvoudigde schema werd de samenvattingsoopdracht van het VWO-examen uit 1982 gebruikt. Uit een groot aantal samenvattingen van examenkandidaten werden 27 samenvattingen van verschillende kwaliteit geselecteerd. Daaraan werden 3 door de onderzoekers zelf geconstrueerde samenvattingen toegevoegd. Volgens de onderzoekers zouden twee van die samenvattingen als 'slecht' en één samenvatting als 'goed' beoordeeld moeten worden. De 30 samenvattingen werden getypt, terwijl namen en dergelijke verwijderd werden.

De 30 samenvattingen werden door 8 docenten Nederlands uit het VWO allereerst beoordeeld aan de hand van de volgende opdracht: "Wilt u de samenvatting als geheel beoordelen, en wel onmiddellijk, zonder analyse van het werkstuk in aspecten. Het is niet de bedoeling dat u gebruik maakt van het in 1982 door de 'commissie vaststelling opgaven' verspreide tekststructuurschema. Na één- of tweemaal goed doorlezen geeft u uw cijfer (schaal 1 t/m 10)". Ongeveer zes maanden na de globale beoordeling van de samenvattingen werden dezelfde samenvattingen door dezelfde 8 beoordelaars met een analytisch beoordelingsschema beoordeeld. Ter controle van de mogelijkheid dat de verwachte winst in overeenstemming met het analytisch schema veroorzaakt was door of de eerder uitgevoerde globale beoordeling of de keuze van de beoordelaars, werden de samenvattingen nog eens door 10 andere docenten beoordeeld.

Van zowel de globale beoordeling als de analytische beoordeling werd de beoordelaarsovereenstemming berekend. De beoordelaarsovereenstemming wordt gedefinieerd als een ratio van variantiecomponenten die geschat worden via een twee-weg variantie-analyse. De maat geeft aan in hoeverre de beoordelingen van beoordelaars identiek zijn. In Sanders et al. (1984) wordt de analyse uitvoerig beschreven.

7. Conclusies, discussie en aanbevelingen

Het gebruik van een analytisch beoordelingsschema bij de beoordeling van de samenvatting Nederlands blijkt tot een hogere overeenstemming tussen beoordelaars te leiden dan de globale beoordeling. De geconstateerde winst in overeenstemming houdt in dat met een globale, onafhankelijke beoordeling door drie beoordelaars dezelfde overeenstemming bereikt wordt als met

een analytische beoordeling door twee onafhankelijke beoordelaars. Deze winst is voldoende reden om bij het eindexamen een analytisch beoordelings-schema in te voeren. Uit de vergelijking van de cijfers van de globale beoordeling met die van de analytische beoordeling blijkt ook in dit onderzoek weer dat kandidaten bij een analytische beoordeling meestal een hoger cijfer ontvangen. Het verschil in totaal gemiddelde tussen globale en analytische beoordeling bedraagt meer dan één cijferpunt.

Hierboven is bij het vermelden van de kenmerken van het nieuwe schema er op gewezen dat de bij het onderzoek gebruikte examentekst uiteen viel in zes beoordelingseenheden (zie bijlage II). Deze beoordelingseenheden bestaan uit bepaalde gedeelten van de tekst, zodat deze voor het gemak ook tekstseenheden genoemd kunnen worden. Om na te gaan of de mate van overeenstemming samenhangt met de tekstlengte, is deze zowel bij de zes onderscheiden tekstseenheden afzonderlijk als bij de eerste drie tekstseenheden samen onderzocht. De analyse van de afzonderlijke tekstseenheden wijst er op dat de omvang van de tekst(eenheden) niet samenhangt met de mate van overeenstemming tussen beoordelaars. Van meer invloed op de overeenstemming tussen beoordelaars lijken twee andere factoren te zijn: de wijze waarop de informatie in de tekst gestructureerd is en de mate waarin deze informatie ondubbelzinnig geïnterpreteerd kan worden.

Het toekennen van aftrekpunten per tekstseenheid voor overbodige informatie levert geen winst in overeenstemming op. In plaats daarvan is het aan te bevelen de huidige regeling van aftrekpunten bij overschrijding van de voorgeschreven lengte van de totale samenvatting te handhaven.

Een bezwaar tegen zowel de globale als de analytische beoordeling zou kunnen zijn dat de beide beoordelingsmethoden niet volledig dezelfde aspecten van de samenvatting beoordelen. Hoewel verder onderzoek daarnaar noodzakelijk is, zijn er aanwijzingen dat de twee methoden hetzelfde meten. De correlatie tussen de gemiddelde cijfers die zijn toegekend door dezelfde 8 beoordelaars die in eerste instantie globaal en vervolgens analytisch beoordeeld hebben is .82. Indien met beide methoden niet hetzelfde gemeten zou worden, zou men in de examensituatie kunnen overwegen één beoordeelaar globaal en een andere beoordeelaar analytisch te laten beoordelen.

Een bezwaar tegen het gebruik van een analytisch beoordelings-schema zou kunnen zijn de omstandigheid dat de beoordelingstijd per samenvatting verlengd wordt. Uit de ervaringen van de beoordelaars bleek echter dat reeds na een korte inwerkperiode snel en efficiënt met een analytisch schema gewerkt kon worden.

Teneinde na te kunnen gaan of de beoordeelaarsovereenstemming die met het analytisch beoordelings-schema gerealiseerd is niet tekstgebonden is, zijn replicaties van het hier beschreven onderzoek met andere teksten noodzakelijk. Verder zal toekomstig onderzoek zich moeten richten op mogelijke verbeteringen van het analytisch beoordelings-schema.

Samenvattingsopdracht

Literatuur

- Drop, W. 1983 *Instrumentele tekstanalyse ten dienste van samenvattingen, opstellen van begripsvragen en tekstverbeteren*. Groningen: Wolters-Noordhoff.
- Drop W, en J.H.L. de Vries 1980 *Taalbeheersing; handboek voor taalhantering*. Groningen: Wolters-Noordhoff. Dijk, T.A. van
- 1978 *Tekswetenschap: een interdisciplinaire inleiding*. Utrecht: Het Spectrum.
- Eemeren, F.H. van et al. 1975 'Samenvatting'. *Nederlands* 2, 41-59.
- Kreeft, H. en T. Luijten 1980 *Verslag van een enquête onder docenten Nederlands Mavo-Havo-VWO naar aanleiding van het C.S.E. tekstbegrip*. Arnhem: Cito (Documentatieroeks 34).
- Rooker, C. 1984 'Over het corrigeren van de samenvatting-VWO'. *Levende talen* 392, 305-312.
- Sanders, P.F., A.C. Hendrix en A.J.M. Luijten 1984 *De beoordeling van de samenvatting Nederlands*. Arnhem: Cito (Specialistisch Bulletin 29).

BIJLAGE I

O-SCHAAL

	wegings-factor
1. structuur inzichtelijk - onduidelijk	5 4 3 2 1
2. Presentatie inzichtelijk - onduidelijk	5 4 3 2 1

F-SCHAAL

	wegings-factor
1. correct taalgebruik - incorrect taalgebruik	5 4 3 2 1
2. eigen formulering - tekstformulering	5 4 3 2 1
3. omslachtige formulering - telegramstijl	5 4 3 2 1

BIJLAGE I (vervolg)

	MAX
Analytisch beoordelingschema bij de samenvatting van de tekst "TOEVAL EN TOELEG IN DE SAMENLEVING" (De Valk)	5
Voorstelling van een zich zelfstandig ontwikkelende maatschappij	1
ervaring van individuen	1
van o.a. de "groten der aarde"	1
voorstelling - onontkoombare conclusie	1
neerslag hiervan in talrijke theorieën (19e eeuwse, geen invloed meer m.u.v. het marxisme)	1
hedendaagse t.: frans structuralisme	1
inhoud: cultuur en geschiedenis bepaald door anonieme structuren	1
eenzelfde ervaring bij hervormers (studenten na 1968)	1
Tegengesteld inzicht: de maatschappij is een menselijk maaksel	5
veranderbaar door groepen met steun van deskundigen/leidersfiguren	1
extreme vertegenwoordigers hiervan: aanhangers v/d samenzweringstheorie	1
inhoud: beheersing van de samenleving door onzichtbare machten	1
Keuze uit beide visies moeilijk vanwege kernen van waarheid in beide visies	5
mogelijkheid van verzoening tussen beide visies	1
ondanks het feit dat het gelijk in beginsel aan de zijde van de aanhangers van de maakbaarheidstheorie is	5
Restrictie: beperkende factoren van natuurlijke aard en van sociale aard (sociale aard:)	5
- vaste structuren	2
noodzakelijke hulpmiddelen voor samenleven, maar verstarrende werking	1
- mens als produkt van de samenleving	2
gevolg: ervaring van de samenleving als natuuriijk gegeven	1
- innerlijke logica van de gegroeide sociale structuur	2
- vervlechting van instellingen	2
gevolg: verandering van a konsekwenties voor b, c etc.	1
gevolg: gebrek aan initiatief	1
- onoverzichtelijkheid samenleving	2
gevolg: geen zicht op konsekwenties van acties	1
- mens als gewoontewezen	2
gevolg: betrekkelijke onveranderlijkheid in doen en laten	1
gevolg: langdurigheid van veranderingen	1
Algemene conclusie: onophoudelijke verandering van de maatschappij, wel door toedoen van mensen, zelden door toeleg	5

totaalscore INHOUD: 60

BIJLAGE II

regel	nr.	Prof. mr. A. Kleijn, De lange en de korte termijn Beoordelingschema	1982-I	score
1	1	De opvatting "regeren is vooruitzien" is algemeen Wel bestaat er verschil in opvatting over de termijn	I	7
	2			2
	3			
7	4	De theorie is duidelijk; er zijn drie termijnen te onderscheiden: lang, middellang en kort	II	7
16	5			7
	6	Maar er is gebrek aan overeenkomst tussen theorie en praktijk Niet gekozen bestuurders benaderen de theorie het meest; korte en middellange termijn voor hen afgeleide van de lange termijn //		7
	7			2
	8	het zwaartepunt berust op de lange termijn		2
	9			2
	9	Gekozen bestuurders hebben voorkeur voor de middellange termijn		2
	10			2
	11	Redenen daarvoor zijn 1) hun beperkte ambtstermijn en 2) hun afhankelijkheid van de kiezersgunst		2
	12			2
	13	Politici en/of volksvertegenwoordigers hebben voorkeur voor korte termijn		2
	14			2
	14	Reden daarvan is dat ze snel politiek succes willen boeken Hierom te weinig aandacht voor de lange termijn // te weinig bekommernis voor het nageslacht		2
	15			2
79	16	Bezinning is noodzakelijk en verloopt moeizaam Vanwege het bestaan van stereotiepe denkpatronen		7
82	17			7
	18	kenmerken hiervan: - onbewust en dwangmatig		2
	19			2
	20	- vaag en onduidelijk		2
	21			2
	21	- algemeen verspreid		2
	22			3
	22	- conformistisch in positieve en negatieve zin		3
	23			3
114	23	Er is behoefte aan non-conformistisch denkwerk op lange termijn		7
	24			2
	24	In ons land zijn alle drie de plantypen vertegenwoordigd		2
	25			2
	26	Dat is een goede zaak		2
	27			2
	27	Vanwege de noodzaak van een visie op lange termijn		2
	28			2
142	28	Conclusie auteur: we moeten plaats inruimen voor non-conformisten in beleidsvorming		7
	29			2
	29	Op een belangrijke en toch voor de democratie ongevaarlijke plaats		2
	30			2
	30	Want zo'n instelling is een onmisbaar element ten dienste van de democratische besluitvorming		2
	31			2

N.B. // geeft aan dat één van beide elementen goed gerekend mag worden.