

---

# Het Effect van Onbewust Redeneren op het Kwantitatief Schatten

---

**Bart Riet, 3455491**  
**Francijn Ottema, 3328724**  
**Lieke Buiks, 3444880**

**Cursus:** Bachelorthesis Onderwijskunde (200600207)  
**Groep:** 01  
**Begeleider:** Jos Jaspers  
**Datum:** 17 juni 2011

*Samenvatting*

In dit onderzoekverslag wordt beschreven in hoeverre er betere antwoorden gegeven worden op kwantitatieve schattingsvraagstukken als deze via onbewust redeneren tot stand komen, dan als deze via bewust redeneren tot stand komen. Tevens wordt onderzocht in hoeverre het toevoegen van afbeeldingen hierop van invloed is. Dit wordt gedaan door middel van een factorieel 2x2 design. Deelnemers aan het onderzoek zijn eerste- en tweedejaars studenten uit het MBO en kinderen uit de twee hoogste groepen van het basisonderwijs. Deze deelnemers zijn aselekt over de vier onderzoekscondities verdeeld. Het onbewust redeneren wordt gestimuleerd door te werken met een distractietaak. Op basis van de literatuur wordt verwacht dat er een hogere kwantitatieve schattingvaardigheidsscore behaald zal worden door het toevoegen van afbeeldingen en door het onbewust redeneren. De uitslagen van ons onderzoek zijn verrassend te noemen, zij gaan namelijk in tegen de verwachte hypothese. Er is geen verbetering waar te nemen op de score van de schattingstest door het toevoegen van afbeeldingen en het werken met een distractietaak. De eerste conditie, waarbij er gewerkt is met een distractietaak en met afbeeldingen, was er zelfs een significant gemiddeld lagere score dan de laatste conditie waarbij er geen afbeeldingen en ook geen distractietaak werd gebruikt. Dit effect trad zowel op in het basisonderwijs als bij de Mbo-studenten. Mbo-studenten scoorden gemiddeld iets beter in het schatten dan de basisschoolleerlingen. Het toevoegen van afbeeldingen heeft wel bijgedragen aan het volledig invullen van de test.

*Sleutelwoorden:* kwantitatief schatten, afbeeldingen, bewust en onbewust redeneren, UTT

### Het Effect van het Onbewust Redeneren op het Kwantitatief Schatten

Het maken van kwantitatieve schattingen is een activiteit die we regelmatig in het dagelijks leven tegenkomen. Denk aan vraagstukken als: Wat kost een herenhuis in het centrum van Utrecht? Hoeveel geld moet ik meenemen op vakantie? Hoe lang doe ik erover om mijn brood te smeren? Bij schatten wordt er geprobeerd adequate antwoorden te geven op vragen waarvoor geen specifiek antwoord beschikbaar is (Burger, 2010). Het kwantitatief schatten is een proces waarbij er numerieke waardes toegeschreven worden aan objecten, gebeurtenissen of abstracties (Brown & Siegler, 1993).

Over het maken van schattingen is inmiddels in verschillende studies onderzoek gedaan. Burger (2010) heeft in een onderwijspsychologische studie onder andere de invloed van de factoren tijdslimiet en persoonlijke karakteristieken op het maken van rekenvaardige en kwantitatieve schattingen onderzocht. Hierbij wordt het rekenvaardig schatten omschreven als het vinden van een passend antwoord op rekenkundige problemen zonder het antwoord precies uit te rekenen (Lemaire, Arnaud & Lecacheur, 2004).

Uit de analyses van Burger blijkt er een zwakke positieve samenhang te bestaan tussen rekenvaardig schatten en kwantitatief schatten. Een tijdslimiet heeft een significante en negatieve invloed op rekenvaardig schatten. Bij kwantitatief schatten was deze invloed echter niet significant.

Geslacht, onderwijsniveau en leeftijd hebben als persoonlijke kenmerken invloed op kwantitatieve schattingen. In een onderzoek van Della Sala, MacPherson, Philips, Sacco en Spinnler (2003) scoorden vrouwen lager dan de mannen op de CET (Cognitive Estimation Task). Respondenten met een lager onderwijsniveau scoorden lager dan respondenten met een hoger niveau. Leeftijd had als enige kenmerk geen invloed op kwantitatief schatten. Wel is hier in de discussie een opmerking gemaakt of de test passend voor iedereen was.

Burger (2010) pleit in de conclusie van haar onderzoek voor nader onderzoek naar de strategieën van kwantitatief schatten en voor het verder verbeteren van de betrouwbaarheid van de gebruikte test. In haar onderzoek heeft zij alleen gebruik gemaakt van proefpersonen welke zestien jaar of ouder zijn. In het onderzoek dat wij voorstaan zullen we naast Mbo-studenten, die allemaal zestien jaar geweest zijn, ook basisschoolleerlingen in de leeftijd van tien tot en met dertien jaar gebruiken. Er zou dus een verschil in leeftijd of schoolsoort waargenomen kunnen worden, aangezien andere onderzoeken hier niet eenduidig over zijn.

Dijksterhuis en Nordgren (2006) beschrijven de nog vrij recente ontdekkingen over de werking van de unconscious-thought theory (UTT). Deze theorie omvat gedachten of

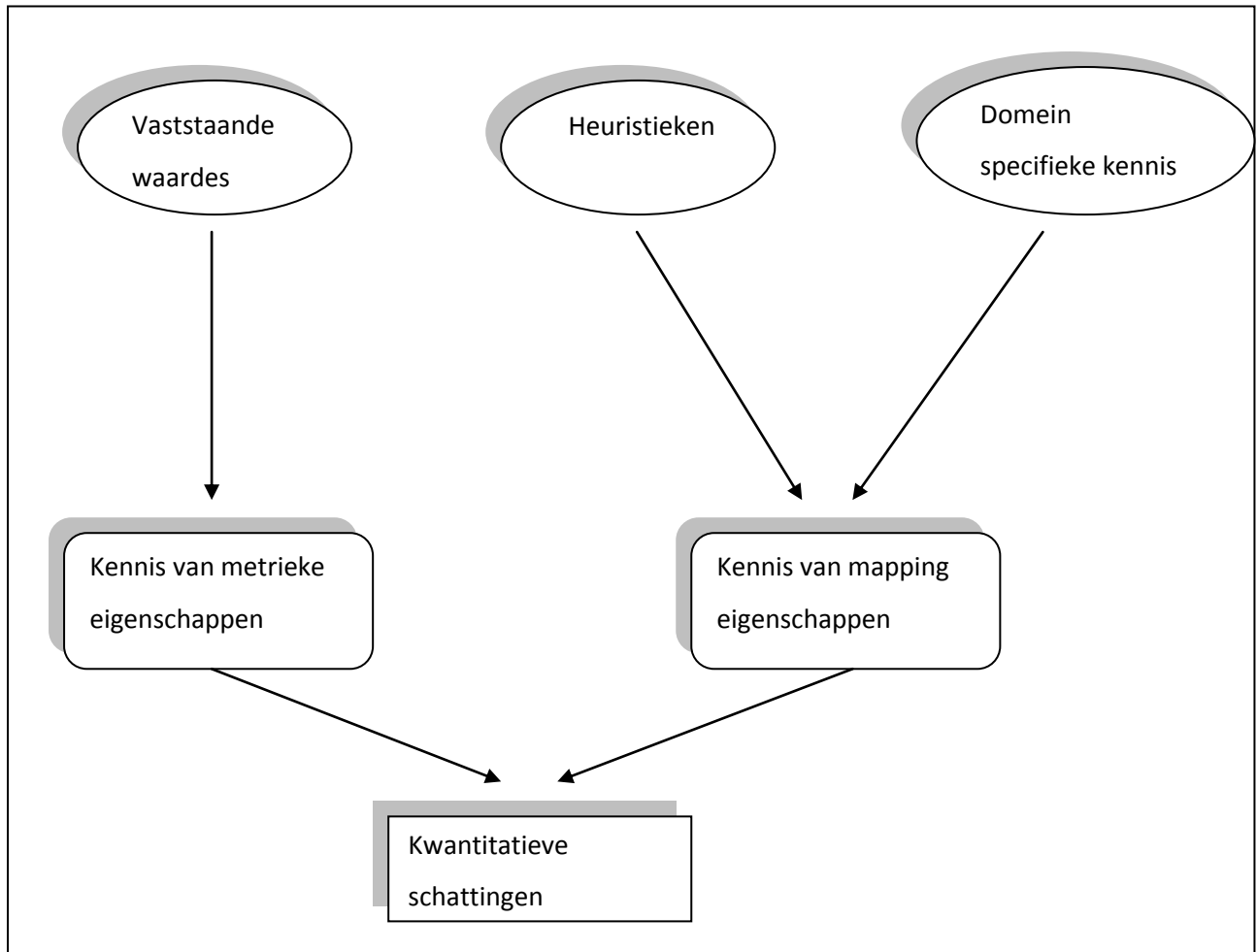
overwegingen zonder bewust aandacht te geven aan het probleem. Het UTT zou effect kunnen hebben op het maken van kwantitatieve schattingen. Daarom is in dit onderzoek onderzocht of het onbewust redeneren inderdaad invloed heeft op het kwantitatief schatten. Aangezien er in een voorlopig meta-onderzoek naar het UTT-effect verwacht wordt dat de werking van het UTT sterker zal zijn naarmate er meer gebruik wordt gemaakt van afbeeldingen, zullen we deze voorspelling meenemen in onze onderzoeksopzet (Strick et al., 2010).

### *Schatten*

Bij een schatting kan niet altijd het precieze antwoord gegeven worden. Er is sprake van een range van adequate antwoorden. Maar er zijn wel strategieën die gebruikt kunnen worden om deze vraagstukken te beantwoorden (Paritosh & Klenk, 2006). Ten eerste, de domein-specifieke kennis in de vorm van regels en/of voorbeelden. Bijvoorbeeld het effect van de huizenmarkt op de prijzen van koopwoningen. Hier worden specifieke en relevante aspecten van het domein aangehaald om tot een juiste schatting te komen. Ten tweede, het gebruik van heuristieken. Dit zijn elementaire regels, vuistregels, die helpen om complexe situaties het hoofd te kunnen bieden. Het zal niet direct het correcte antwoord opleveren, maar wel een zuiverdere schatting kunnen geven (Zimbardo, Johnson & McCann, 2009). Deze methodische weg is een handvat die een deel van de oplossing kan bieden. Men kiest dan voor de meest gebruikte heuristiek. Bijvoorbeeld wanneer het aantal scheidingen in een stad geschat wordt aan de hand van het aantal scheidingen dat men in de eigen omgeving kent. De laatste strategie die gebruikt kan worden om schattingen te maken, zijn vaststaande waardes. Om bijvoorbeeld het vriespunt van wodka te schatten kan er gebruik gemaakt worden van het vriespunt van water (Brown & Siegler, 1993). Dit zijn allemaal voorbeelden van informatierijke en alledaagse schattingen. In dit soort situaties wordt gebruik gemaakt van analogische schattingen. Dit zijn voorbeelden uit eerdere ervaringen (Paritosh & Klenk, 2006).

In Figuur 1 is te zien hoe de drie strategieën gebruikt kunnen worden om hedendaagse kwantitatieve schattingen te kunnen maken. Deze strategieën zijn weer onderverdeeld in de twee componenten; metrieke- en mappingeigenschappen. Metrieke kennis richt zich op statistische eigenschappen, zoals het gemiddelde, mediaan en varianties. Mapping richt zich op de relaties tussen individuele entiteiten binnen het domein. Metriek en mapping samen

geven specifieke informatie over datgene wat geschat moet worden, echter apart van elkaar doen ze dat niet (Brown & Siegler, 1993).



*Figuur 1.* Basisstructuur van metrieke - en mapping eigenschappen bij kwantitatieve schattingen.

Toch worden de strategieën niet altijd goed gebruikt. Bij heuristieken worden vuistregels ingezet om via een methodische weg naar een schatting te komen. Maar als men de verkeerde vuistregel hanteert, dan klopt de schatting ook niet. Men kent misschien meer scheidingen in de directe omgeving, dan het aantal dat gebruikt wordt om een schatting te maken. Als de vergelijkende waarde niet klopt, dan kan er ook geen goede schatting gemaakt worden (Tversky & Kahneman, 1974).

Tijdsdruk is een factor die invloed heeft op de uitkomsten bij kwantitatief schatten. Uit onderzoek van Rhymer en collega's (2002) blijkt dat studenten die voorgelegde opgaven onder tijdsdruk moesten maken, meer opgaven maakten en meer correcte antwoorden gaven

dan de studenten die niet onder tijdsdruk werkten. Tijdsbeperking om een bepaalde taak af te ronden blijkt in de praktijk verschillende effecten te hebben die afhangen van de complexiteit van de taak. De tijdsdruk bleek vooral effectief te zijn voor eenvoudige rekentaken. Voor complexe rekentaken is het niet raadzaam om de studenten onder tijdsdruk te laten werken omdat dit de resultaten negatief beïnvloedt. Één van de conclusies in het onderzoek van Bos, Dijksterhuis en van Baaren (2011) is dat het onbewuste denken vooral bij de levenshouding van de huidige ouders en grootouders past om hun beslissingen niet overhaast te maken.

### *Het onbewuste*

Onderzoeken naar de werking van het onbewuste, zijn al meer dan een eeuw oud. Grondlegger van de psychoanalyse Sigmund Freud introduceerde samen met Josef Breuer in de klassieker '*Studien über Hysterie*' het begrip '*Unbewusste*' (Breuer & Freud, 1896). Dijksterhuis en Nordgren (2006) introduceren meer dan een eeuw later *de unconscious-thought theory (UTT)*. Zij onderscheiden een bewuste gedachtegang en een gedachtegang waarbij onbewust geredeneerd wordt. Afhankelijk van de complexiteit van de uit te voeren taken is een voorkeur voor een strategie aan te wijzen. Eenvoudige vraagstukken kunnen het beste via bewust redeneren opgelost worden. Complexe taken, waarbij meerdere aspecten nodig zijn om de taak op te lossen, kunnen echter beter zonder bewust redeneren worden opgelost. De theorie gaat ervan uit dat de hersenen via de onbewuste gedachtegang ervoor zorgen dat er toch adequatere keuzes gemaakt worden. Dit wordt ook wel het '*Deliberation-without-attention effect*' genoemd (Dijksterhuis, Bos, Nordgren & van Baaren, 2006).

Aart en Hassin (2005) hebben aangetoond dat het onbewuste beslag kan leggen op de capaciteit van het werkgeheugen. Het onbewuste kan op diverse manieren geactiveerd worden. De sociale omgeving is een belangrijke stimulator om het onbewuste tot doelgericht gedrag te activeren (Aarts et al., 2005). Daarnaast spelen ook de zintuigen een rol bij het activeren van het onbewuste (Aarts, 2006). In het artikel van Nordgren, Bos en Dijksterhuis (2011) blijkt dat een combinatie van zowel bewust als onbewust schatten tot de beste keuzes leidt. De meest optimale keuzes worden verkregen wanneer het onbewust denken het bewust denken opvolgt.

De uitkomsten van het onderzoek naar UTT zijn verrassend te noemen. Het gaat tegen de eigen intuïtie in. De publicaties van dit onderzoek in het gerenommeerde tijdschrift '*Science*' heeft onderzoekers aangezet om deze theorie nader te onderzoeken. Een eerste

meta-analyse over UTT lijkt inderdaad te bevestigen dat de kans op het UTT het grootst is bij complexe keuzes (Strick et al., 2010).

Er is ook kritiek op het UTT. González-Vallejo en Phillips (2010) wijzen erop dat er nog tal van onduidelijkheden en tegenstrijdigheden zijn. Een interessante onduidelijkheid is dat het onbewuste volgens Dijksterhuis en Nordgren (2006) niet bij het uitrekenen van sommen zou werken. De verklaring die hierbij gegeven wordt, is dat het antwoord via een vast regelprincipe verloopt. Het onbewuste kan zich wel conformeren naar regels, maar ze niet strikt volgen. Het onbewuste is wel in staat om getallen toe te wijzen aan hoeveelheden, maar dit zullen dan ruwe schattingen zijn (Betsch, Plessner, Schwieren & Gütig, 2001). Lassiter, Lindberg, Gonzalez-Vallejo, Bellezza en Phillips (2009) beredeneren dat er geen onbewust denken bestaat. Zij noemen het UTT een eerste overweging uit het geheugen.

In dit onderzoek wordt ook de aanwezigheid van afbeeldingen bij de schattingsvragen betrokken. Het is al bekend dat afbeeldingen als toegang kunnen dienen voor het impliciete geheugen. Dit wordt ook wel priming genoemd (Gabrieli, Fleischman, Keane, Reminger, & Morell, 1995). Daarnaast zal in een voorlopig meta-onderzoek naar het UTT-effect worden verwacht dat de werking van het UTT sterker zal zijn wanneer er gebruik wordt gemaakt van afbeeldingen (Strick et al., 2010). Er is nog niet eerder onderzocht in hoeverre het toevoegen van afbeeldingen aan kwantitatieve schattingsvragen meer of minder effect zal hebben bij bewust redeneren versus onbewust redeneren.

### *Vraagstelling*

De hoofdvraag van dit onderzoek is: *in hoeverre worden er betere antwoorden op kwantitatieve schattingsvraagstukken gegeven als deze via onbewust redeneren tot stand komen dan als ze via bewust redeneren tot stand komen?* Hierbij zal tevens onderzocht worden of dit effect afhankelijk is van de aanwezigheid van afbeeldingen bij schattingsvragen.

Gekoppeld aan deze hoofdvraag zijn een aantal deelvragen die mede door de literatuur gevormd zijn. Ten eerste: *is er verschil tussen de kwantitatieve schattingsvaardigheid van eerste - en tweedejaars Mbo-studenten (zestien tot en met achttien jaar) en de basisschoolleerlingen (tien tot en met dertien jaar)?*

Ten tweede: *in hoeverre is sekse van invloed op het kwantitatief schatten?* Ten derde *in hoeverre is leeftijd en onderwijsniveau van invloed op het kwantitatief schatten?* Leeftijd en onderwijsniveau zijn bij deze deelvraag gekoppeld. Als er verschillen zijn tussen Mbo-

studenten en basisschool leerlingen, dan zou dat immers zowel verklaard kunnen worden door leeftijd als door middel van verschil in onderwijsniveau.

Op basis van de aangehaalde literatuur verwachten we een hogere score op de Biber-schattingstest wanneer er gebruik wordt gemaakt van een distractietaak om het onbewust redeneren te stimuleren. Door het toevoegen van afbeeldingen verwachten we dat de score op de Biber-schattingstest nog hoger zal worden. Op basis van het onderzoek van Della Salla et al. (2003) verwachten we dat vrouwelijke deelnemers iets beter zijn in schatten dan de mannelijke deelnemers.

Als laatste wordt dit onderzoek gebruikt om aanbevelingen te doen om de Biber-schattingstest voor soortgelijk onderzoek te kunnen verbeteren.

### **Methode**

#### **Design**

In het onderzoek is er sprake van een 2x2 factorieel onderzoeksdesign. De deelnemers zijn aselekt aan de vier verschillende condities toegewezen. De afhankelijke variabele is de vaardigheidsscore bij het kwantitatief schatten. De eerste onafhankelijke variabele is het bewust redeneren versus onbewust redeneren. De tweede onafhankelijke variabele is de aanwezigheid van afbeeldingen bij de schattingsvragen. Het onbewust redeneren werd gestimuleerd door te werken met een distractietaak. Deze distractietaak was overeenkomstig de aanbevelingen van Strick et al. (2010) een woordzoeker. De distractietaak werd bij bewust redeneren niet gebruikt. Zo ontstonden uiteindelijk de volgende vier testcondities.

Test 1: met distractie en toegevoegde afbeeldingen

Test 2: met distractie maar zonder afbeeldingen

Test 3: zonder distractie maar met toegevoegde afbeeldingen

Test 4: zonder distractie en zonder afbeeldingen

In totaal zijn er 375 testen gemaakt, waarvan er 220 vragenlijsten volledig doorlopen zijn. Er werden totaal 155 testen voortijdig afgebroken. Deze testen zijn niet meegenomen bij de analyses. De verdeling van het aantal respondenten was vrijwel gelijk. Een frequentieoverzicht van het aantal afgenomen vragenlijsten en het percentage vragenlijsten dat volledig doorlopen is, is weergegeven in Tabel 1.



Tabel 1

*Overzicht van het Aantal Vragenlijsten dat is Uitgezet en Kolompercentages al dan niet Volledig Ingevulde Vragenlijsten (totale N=375)*

Respons	Aantallen				Percentages			
	Test 1	Test 2	Test 3	Test 4	Test 1	Test 2	Test 3	Test 4
Volledige reacties	55	50	55	60	56	50	69	63
Onvolledige reacties	44	50	25	36	44	50	21	37
Totaal aantal reacties	99	100	80	96				

### Deelnemers

Er is getracht om een zo groot mogelijke steekproef te werven binnen de eerste en tweede klas van de Mbo-opleiding Welzijn binnen het Hoornbeek College te Amersfoort. 132 Mbo-studenten hebben de test volledig gemaakt. De studenten hebben leeftijden tussen de 16 en 18 jaar. Daarnaast is de test afgenomen in de groepen zeven en acht van het basisonderwijs. Deze leerlingen hebben leeftijden tussen de 10 en 13 jaar. Basisschool de Startbaan te Eindhoven en basisschool de Brug in Harderwijk hebben hiervoor aan dit onderzoek hun medewerking verleend. Hier lukte het 88 leerlingen om de test volledig te doorlopen. De deelnemers zijn aselekt aan de 4 testcondities toegewezen. In Tabel 2 is te zien hoe de geslachten over de opleidingen en de vier verschillende testcondities verdeeld zijn.

Tabel 2

*Frequentieoverzicht van Aantal Respondenten Uitgesplitst naar Geslacht en Opleiding over de vier Testcondities (totale N=220)*

	Vrouwen		Mannen	
	Basonderwijs	MBO	Basonderwijs	MBO
Test 1 plaatje + distractie	12	33	10	0
Test 2 zonder plaatjes + distractie	8	26	13	3
Test 3 plaatjes zonder distractie	13	30	11	1
Test 4 zonder plaatsjes zonder distractie	6	38	15	1
Totaal	39	127	49	5

Zoals in Tabel 2 te zien is, waren de mannelijke respondenten op het MBO ondervertegenwoordigd. Dit is te verklaren doordat de richting Welzijn voornamelijk vrouwelijke studenten heeft. In het basisonderwijs is de verhouding man-vrouw meer in evenwicht en zijn er zelfs iets meer mannen.

### **Instrumenten**

Voor het meten van de schattingsvaardigheid is gebruik gemaakt van de Biber-schattingstest (Bullard et al., 2004). Deze test is in het Nederlands vertaald door Burger (2010). De test is ontworpen en uitgevoerd met behulp van het computerprogramma Limesurvey. Dit programma kan de tijd beheersen en er kunnen een onbeperkt aantal vragenlijsten op hetzelfde moment anoniem worden afgenomen. We hebben in het programma Limesurvey vier verschillende soorten vragenlijsten ontworpen, welke geschikt zijn gemaakt voor de al eerder genoemde vier verschillende testcondities. De distractietaak die bij twee van de vier testcondities voorkomt is een woordzoeker. Woordzoekers zijn een goede distractietaak voor het onbewust redeneren gebleken (Strick et al., 2010).

In dit onderzoek wordt gekeken of de Nederlandse versie van de Biber-schattingstest een betrouwbare vragenlijst is om het kwantitatief schatten te meten. In het onderzoek van Bullard (2004) heeft deze test een goede betrouwbaarheid (Cronbach's alpha is .62). In ons onderzoek maken we gebruik van de vertaling van de vragenlijst in het Nederlands door Burger (2010). Tevens scoren we de vragen, overeenkomstig de aanbevelingen van Burger volgens de manier van Axelrod en Mellis (1994). Dit wordt gedaan om het onderscheidend vermogen van de test te vergroten. Antwoorden onder het 2<sup>de</sup> percentiel, of boven het 98<sup>ste</sup> percentiel krijgen hierbij nul punten. Antwoorden tussen het 2<sup>de</sup> percentiel en 16<sup>de</sup> percentiel en antwoorden tussen de 84<sup>ste</sup> percentiel en 98<sup>ste</sup> percentiel krijgen hierbij één punt. De antwoorden tussen percentiel 16 en 84 scoren twee punten. Voor het bepalen van de goede antwoorden is net als in het onderzoek van Bullard (2010) de door de test zelf gegenereerde data gebruikt. In Tabel 3 is een overzicht te zien van de gehanteerde schattingsvragen tevens zijn in Tabel 3 de waarden van de percentielen af te lezen waartussen de punten worden toegekend.

Tabel 3

*Normeringtabel met Schattingsvragen en Percentielen waartussen Punten worden Toegekend*

Vraag	2 <sup>de</sup>	16 <sup>de</sup>	84 <sup>ste</sup>	98 <sup>ste</sup>
1. Hoeveel pitten zitten er in een watermeloen?	2,26	30,00	500,00	4200,00
2. Hoeveel weegt een mobiele telefoon?	1,24	25,00	500,00	1380,00
3. Hoeveel stokjes spaghetti zitten er in een pak van één pond?	13,20	50,00	500,00	4032,00
4. Wat is de afstand dat een volwassene in een avond kan lopen?	2,55	10,00	40,00	100,00
5. Hoe hoog kan iemand op een trampoline springen?	1,00	1,50	4,00	12,7,00
6. Hoe lang duurt het bouwen van een woonhuis?	2,88	25,10	365,00	824,00
7. Hoeveel wegen 12 normale appels bij elkaar?	0,50	1,20	50,00	2420,00
8. Wat is de afstand dat een paard en wagen in één uur kan afleggen?	2,00	8,00	40,00	100,00
9. Hoe vaak kan je je tanden poetsen van een tube tandpasta?	6,72	20,00	100,00	475,70
10. Hoeveel chips zitten er in een gewone zak van 200 gram?	12,30	45,00	367,40	2460,00
11. Hoelang duurt het voor een volwassene om één kant van een A4 vol te schrijven?	1,50	4,00	20,00	102,00
12. Wat is de leeftijd van de oudste, levende persoon in Nederland?	90,60	102,00	115,00	147,70
13. Hoe lang is een eetlepel?	2,28	10,00	25,00	50,00
14. Hoeveel weegt een inklapbare tuinstoel?	0,50	2,00	30,00	2910,00
15. Hoeveel tijd kost het om een T-shirt te strijken?	0,55	2,00	5,00	30,00
16. Hoe lang is de nek van een giraffe?	1,00	1,50	4,00	18,86
17. Hoeveel plakken brood zitten er in een brood van 500 gram?	3,32	10,00	25,00	53,40
18. Hoeveel weegt een paar mannenschoenen?	1,00	5,00	900,00	3000,00
19. Hoe zwaar is een paard?	18,64	75,00	1000,00	30400,00
20. Hoe lang duurt het voordat verse melk na de houdbaarheidsdatum in de koelkast zuur wordt?	1,00	2,00	10,92	70,00

*Noot:* De vragen zijn overgenomen van Burger (2010). Hierbij is gebruik gemaakt van de Biber-schattingstest van Bullard (2004). Vraag 19 was hierbij cultuurafhankelijk en is daarom vervangen door een soortgelijke vraag in dezelfde categorie.

### **Procedure**

De testen in het programma Limesurvey waren zo geprogrammeerd dat er aan het begin van de test eerst naar de geboortedatum, geslacht en het opleidingsniveau van de deelnemers werd gevraagd. De deelnemers waren vooraf aselekt toegewezen aan de vier testcondities. Voordat de test begon kreeg elke leerling een strookje waarop de naam vermeld

stond, de versie van de test die gemaakt moest worden, de link van de URL en een korte instructie. Aan de deelnemers van conditie 1 en 2 werden ook de distractietaak uitgereikt.

De testen zijn per klas in het computerlokaal afgenomen. Bij testcondities 1 en 2 werd elke vraag afzonderlijk vijftien seconden lang op het beeldscherm getoond. Aan testconditie 1 waren afbeeldingen toegevoegd. Het was niet mogelijk om tijdens deze fase van de test al antwoorden in te vullen. Aansluitend kregen beide testcondities een woordzoeker gepresenteerd. De deelnemers kregen deze zowel op papier als op het beeldscherm te zien. Voor het maken van de woordzoeker kregen de deelnemers vier minuten de tijd. Ze werden door de leerkracht aangemoedigd om de woordzoeker goed te maken. Door deze distractietaak konden de antwoorden op de schattingsvragen alleen via het onbewust redeneren bedacht worden. Na de vier minuten mochten de deelnemers niet verder met de distractietaak en moest deze worden weggelegd. Op het beeldscherm verschenen weer alle schattingsvragen afzonderlijk en werd er vijftien seconden de tijd gegeven om elke vraag te beantwoorden.

Bij de testcondities 3 en 4 kregen de deelnemers vijf minuten de tijd om de schattingsvragen te bestuderen. De schattingsvragen werden allemaal tegelijk gepresenteerd, zodat de deelnemers zelf konden bepalen hoe lang ze over een vraag wilden nadenken. Op deze manier werd het bewust redeneren gestimuleerd. Bij testconditie 3 waren dezelfde afbeeldingen uit testconditie 1 aan de schattingsvragen toegevoegd. Het tussentijds invullen van de antwoorden was tijdens deze vier minuten niet mogelijk. Aansluitend verscheen elke vraag nog een keer afzonderlijk op het beeldscherm. De deelnemers kregen ook hier vijftien seconden de tijd om een antwoord in te vullen.

### **Resultaten**

#### *Betrouwbaarheid van de schaal*

Allereerst is de betrouwbaarheid van het instrument bepaald. Deze lijkt aanvankelijk niet hoog te zijn. Cronbach's alfa is 0.327. Na het verwijderen van de vragen: 1, 4, 16, 17, 18 en 20 stijgt Cronbach's alfa tot 0.436. We hebben besloten deze items om inhoudelijke redenen echter niet te verwijderen. De items zijn immers een vast onderdeel van de al bestaande Biber-test. In het artikel van Bullard (2004) had deze test met een Cronbach's alfa van 0.62 een hogere betrouwbaarheid. Voorzichtigheid bij de interpretatie van de testuitslag is gezien de lage alfa daarom echter wel geboden. De lage betrouwbaarheid wordt voor een belangrijk deel echter veroorzaakt doordat alle missing values niet betrokken worden in het

berekenen van de betrouwbaarheid. Als we aan alle missing values nul punten toekennen wordt de betrouwbaarheid van de schaal ineens verhoogd tot een Cronbach's alfa van 0.566. De betrouwbaarheid is hiermee van voldoende omvang. Na het verwijderen van vraag 7, 16, 17, 18, 19 en 20 zou het zelfs mogelijk zijn om de betrouwbaarheid te laten stijgen tot een Cronbach's alfa van 0.648. Om de al eerder genoemde reden wordt echter besloten dit niet te doen. Er is nog bekeken of de betrouwbaarheid hoger zou zijn wanneer deze alleen berekend zou worden voor het MBO of voor de basisschool. Dit bleek voor het MBO inderdaad het zo te zijn. Hier werd Cronbach's alfa 0.642 mits de missing values weer op nul werden gezet. Na verwijdering van vraag 7, 16, 17, 18, 19 en 20 steeg Cronbach's alfa zelfs tot 0.712. Voor alleen het basisonderwijs werd met dezelfde tactiek een betrouwbaarheidsschaal met een Cronbach's alfa verkregen van 0.440. Deze lage betrouwbaarheid wordt voor een belangrijk deel veroorzaakt door de geringe omvang van het aantal respondenten in de basisschool (n=88).

#### *Bewust redeneren versus onbewust redeneren en het gebruik van afbeeldingen*

In Tabel 4 zijn de gemiddelde schattingscores over de vier testcondities weergegeven. Om de hoofdvraag te kunnen beantwoorden wordt er een meervoudige variantieanalyse met de factoren afbeeldingen en distractie toegepast. De afhankelijke variabele is de totale schattingscore. We constateren dat er zowel een significant verschil bestaat in de totale schattingscore bij het al dan niet toevoegen van afbeeldingen ( $F(1,216) = 5,604$ ;  $p=0,019$ ) als bij het al dan niet toevoegen van distractie ( $F(1,216) = 4,129$ ;  $p=0,043$ ). Er is geen significant interactie-effect.

Tabel 4

#### *Beschrijvende Statistiek van de Totale Schattingscore verdeeld over de vier Testcondities*

Totaalscore	Gemiddelde	S.D.	N
Test 1 met plaatjes en met distractie	29,62	5,12	55
Test 2 zonder plaatjes en met distractie	31,28	4,31	50
Test 3 met plaatjes , zonder distractie	31,09	3,67	55
Test 4 zonder plaatsjes, zonder distractie	32,10	3,48	60
Totaal	31,04	4,25	220

Op de testen waarbij afbeeldingen toegevoegd zijn (gemiddeld 30,45; sd 4,49) wordt slechter gescoord dan wanneer er geen afbeeldingen zijn toegevoegd (gemiddeld 31,73; sd 3,88). Ook het gebruiken van distractie (gemiddeld 30,41; sd 4,80) om het onbewust redeneren te stimuleren scoort lager dan wanneer er geen distractie plaatsvindt (gemiddeld 31,62; sd 3,59). Voor beide factoren is de effectgrootte berekend. De aanwezigheid van afbeeldingen kan 2,5 procent van de totale schattingscore verklaren. De aanwezigheid van distractie kan 1,9 procent van de totale schattingscore verklaren. De relevantie voor beide factoren is laag.

Bij de schattingstest had het bewust redeneren een gemiddelde van 31,62 (sd = 3,59). Het onbewust redeneren scoorde een lager gemiddelde van 30,41 (sd = 4,80). Uit een t-toets blijkt dat dit verschil significant is bij tweezijdige toetsing ( $t(1,192)=2,098$  ;  $p=0,037$ ).

#### *Persoonlijke karakteristieken*

Door middel van t-toetsen is de invloed van sekse, leeftijd en onderwijsniveau op kwantitatief schatten bepaald. Vrouwen (n=164) scoren iets beter dan mannen (n=54). Vrouwen scoren gemiddeld 31,38 punten met een standaarddeviatie van 4,43 en mannen scoren gemiddeld 30,00 punten met een standaarddeviatie van 3,43. Dit verschil is significant;  $t = 2,378$ ;  $df = 115,313$ ;  $p=0,019$  bij tweezijdige toetsing. Het geslacht kan 2 procent van de verschillen verklaren. De relevantie is dan ook gering.

Tabel 5

#### *Beschrijvende Statistiek van de Totaalscore verdeeld over de vier Testcondities Uitgesplitst naar Opleiding*

Totaalscore	Basisonderwijs			MBO		
	Gemiddelde	SD	N	Gemiddelde	SD	N
Test 1 met plaatjes en met distractie	29,41	2,56	22	29,76	6,31	33
Test 2 zonder plaatjes, met distractie	30,14	3,37	21	32,10	4,77	29
Test 3 met plaatjes, zonder distractie	28,96	3,67	24	32,74	2,74	31
Test 4 zonder plaatsjes, zonder distractie	30,00	3,48	21	33,23	2,94	39
Totaal	29,60	3,28	88	32,00	4,55	132

Uit Tabel 5 valt op te maken dat de kinderen van het basisonderwijs (gemiddelde score 29,60 ; sd =3,28) gemiddeld genomen slechter scoren in het schatten dan de studenten van het MBO (gemiddelde score 32,00 ; sd =4,55). Dit verschil is significant;  $t = -4,26$ ;  $df=218$ ;

$p < 0,001$  bij tweezijdige toetsing. Schoolsoort kan ongeveer 6% van de verschillen verklaren ( $d = 0,57$ ). Dit komt overeen met een matige effectgrootte.

### Conclusie

De centrale vraag bij dit onderzoek is: *in hoeverre worden er betere antwoorden op kwantitatieve schattingsvraagstukken gegeven als deze via onbewust redeneren tot stand komen dan als ze via bewust redeneren tot stand komen? In hoeverre is de aanwezigheid van afbeeldingen van invloed op de schattingscore.* Aan deze onderzoeksvraag zijn enkele deelvragen als de invloed van de factoren sekse, schoolsoort en leeftijd toegevoegd. Deze kunnen beantwoord worden door de analyses uit de resultaten.

Allereerst is het bewust redeneren vergeleken met het onbewust redeneren. Uit de resultaten blijkt dat het onbewust redeneren een iets lager gemiddelde heeft dan bewust redeneren. Dit verschil bleek ook significant te zijn. Dit is dus juist het tegenovergestelde dan wat de verwachting was. Het komt niet overeen met wat Dijksterhuis en Nordgren (2006) beschrijven in hun UTT (*unconscious-thought theory*).

Het inzetten van een distractietaak heeft bijgedragen om de totaalscore op de schattingsstest te verlagen. Ook het toevoegen van afbeeldingen heeft een negatieve invloed op de totale schattingscore. Wel heeft het geholpen om de testen volledig af te maken. Dat het toevoegen van afbeeldingen een negatieve invloed heeft spreekt het onderzoek van Strick et al. (2010) tegen. Om een zo hoog mogelijke score op de Biber-schattingsstest te behalen kan er beter bewust geredeneerd worden en de afbeeldingen achterwege worden gelaten.

Als laatst werd de invloed van persoonlijke karakteristieken als geslacht, leeftijd en onderwijsniveau onderzocht. Het blijkt dat vrouwen iets hoger scoren dan mannen. Dit is wederom verassend omdat dit niet overeen komt met het onderzoek van Della Sala et al. (2003), waarin vrouwen juist lager scoorden. De basisschoolleerlingen scoorden op alle vier de testcondities lager dan de Mbo-studenten. Leeftijd en onderwijsniveau kan dus van invloed zijn op kwantitatieve schattingsvraagstukken zowel bij het bewust redeneren als het onbewust redeneren. Toch is hier een kanttekening te plaatsen. Het zou ook goed mogelijk zijn dat de test voor basisschool leerlingen te moeilijk is. Dit zal verder in de discussie besproken worden.

De betrouwbaarheid van de test was door de vele missing values aanvankelijk niet erg hoog. De tegenstrijdigheden die zijn gevonden kunnen daarmee verklaard worden. Daarom

kan niet alles zomaar aangenomen worden, zeker als de literatuur het tegendeel bewijst. In de discussie zal hier verder op in worden gegaan.

### **Discussie**

Dit onderzoek heeft enkele verrassende bevindingen gedaan omtrent bewust en onbewust schatten. Wat er in de literatuur is gevonden, kwam niet in alles overeen met de uitkomsten van dit onderzoek. We zullen hieronder proberen te verklaren waarom er zoveel verschillen zijn met andere onderzoeken.

De Biber-schattingstest is aanvankelijk ontwikkeld voor personen met een minimumleeftijd van zestien jaar. Dit kan een oorzaak zijn dat niet alleen de basisschoolleerlingen moeite met de opzet van de test hadden, maar ook de Mbo-studenten. Dit wordt ondersteund doordat de betrouwbaarheid van het instrument bij alleen de Mbo-studenten aanmerkelijk hoger was.

De leerlingen van de basisschool vonden het lastig om de persoonsgegevens in te vullen. De optie 'MBO', 'basisonderwijs' en 'anders' was voor hen erg verwarrend. Dit had voor de duidelijkheid beter 'groep 7', 'groep 8' en 'MBO' kunnen zijn.

Bij twee testen (testconditie drie en vier) kregen de leerlingen vijf minuten de tijd om de vragen in zich op te nemen en alvast antwoorden te bedenken. Dit hebben meerdere leerlingen als te lang ervaren. Ze hadden moeite om geconcentreerd te blijven. Het nut werd er niet van ingezien en de verveling sloeg toe. Dat de betrouwbaarheid van de test aanmerkelijk verbeterd door de laatste 5 vragen te verwijderen is dan ook niet geheel verwonderlijk. Er kan afgevraagd worden in hoeverre de vragen daadwerkelijk bewust zijn gelezen. De moeilijkheidsgraad van de distractietaak was goed. Tijdens de distractietaak gingen sommige leerlingen overleggen over de woordzoeker. Dit gaf enige onrust, maar zou geen invloed moeten hebben op het idee van distractie.

Na de distractietaak of na het bekijken van de vragen, moesten alle leerlingen de schattingsvragen gaan beantwoorden. Veel leerlingen kregen het niet voor elkaar om in de gegeven tijd van vijftien seconden het antwoord in te vullen. Hierdoor zijn mogelijk foutieve antwoorden gegeven of is er helemaal niets ingevuld. Een oorzaak hiervoor zou het niet geconcentreerd bekijken van de vragen tijdens de eerste vijf minuten kunnen zijn. Het viel op dat bij het invullen erg veel leerlingen zich opgejaagd voelden vanwege de tijdsdruk en de vragen bekeken alsof ze deze voor het eerst zagen. Hierdoor hadden ze weinig tijd om na te denken. Dit zou een oorzaak kunnen zijn dat er relatief veel testen niet zijn afgemaakt.

Daarnaast was het verwarrend dat de leerlingen de vragen eerst 10 seconden te zien



kregen en dat daarbij ook de tijd afliep. Ze hadden het idee dat er al meteen iets ingevuld moest worden en dat was niet mogelijk, omdat de vragen eerst alleen bekeken konden worden. Pas na het bekijken of na de distractietaak konden ze antwoorden invullen. Omdat de leerlingen of de uitleg niet goed gelezen hadden, of deze niet goed begrepen hadden, dachten ze dat ze niets in konden vullen en dat de test niet goed werkte. Ze braken hierdoor de test voortijdig af. Een groot aantal leerlingen begon niet nogmaals aan de test en op deze manier ontstonden er veel onafgemaakte testen.

Een aanbeveling voor vervolgonderzoek is de tijd voor het bewust redeneren met één minuut terug te brengen naar vier minuten. De tijd om een vraag te beantwoorden zou voor alle vier de testcondities met vijf seconden verlengd moeten worden naar twintig seconden per vraag. Een andere optie zou kunnen zijn om ook een kort pilotonderzoek te doen onder leerkrachten en leerlingen. Leerkrachten kunnen dan bepalen of ze de test geschikt achten voor hun klas. Door de test eerst uit te proberen bij een groepje leerlingen, kunnen onderzoekers de tijdsduur dat de vragen getoond worden eventueel aanpassen. De tijd die staat voor het doorlezen en invullen van de vragen zal dan misschien voor elke doelgroep anders zijn.

Wat in deze testopzet niet onderzocht is, is in hoeverre intelligentie een rol speelt bij het kwantitatief schatten. Omdat tijdens de observaties het opviel dat de kinderen waarvan bekend is dat zij over een goed cognitief intellect beschikken een meer geconcentreerde aanpak hanteerde, lijkt het de moeite waard om hier ook vervolgonderzoek naar te doen.

De beperkte betrouwbaarheid van de schaal zou mogelijk veroorzaakt kunnen worden doordat Cronbach's alfa berekend is over het totaal van de vier verschillende testen. Het aantal deelnemers per testconditie is echter te laag om te leiden tot een hogere betrouwbaarheid. Het is aan te raden om tijdens vervolgonderzoek te werken met grotere steekproeven welke ook oudere deelnemers van meerdere opleidingsniveaus bevatten.

### Referenties

- Aarts, H. (2006). Onbewust doelgericht gedrag en de corrosie van de ijzeren wil. Rede uitgesproken bij de aanvaarding van het ambt van hoogleraar, Universiteit Utrecht.
- Aarts, H., Chartrand, T.L., Custers, R., Danner, U., Dik, G., Jefferis, V.E., & Cheng, C.M. (2005). Social stereotypes and automatic goal-pursuit. *Social Cognition*, 23, 464-489.
- Aarts, H. & Hassin, R. (2005). Automatic goal inferences and contagion: On pursuing goals one perceives in other people's behavior. In J.P. Forgas, D.D. Williams & W. Von

- Hippel, (Eds.), *Social motivation: Conscious and unconscious processes* (pp. 153-167). New York: Psychology Press.
- Betsch, T., Plessner, H., Schwieren, C., & Gütig, R. (2001). I like it but I don't know why: A value-account approach to implicit attitude formation. *Personality and Social Psychology Bulletin*, 27, 242–253.
- Bos, M.W., Dijksterhuis, A. and van Baaren, R.B. (2011). The benefits of "sleeping on things": Unconscious thought leads to automatic weighting. *Journal of consumer psychology*, 21 (1 ), 4-8
- Breuer, J., & Freud, S. (1896). *Studien über Hysterie*. Leipzig
- Brown, N.R. & Siegler, R.S. (1993). Metrics and Mappings: A framework for understanding real world quantitative estimation. *Psychological review*. 100(3). 511-534.
- Bullard, S.E., Fein, D., Gleeson, M.K., Tischer, N., Mapou, R.L., & Kaplan, E. (2004). The Biber cognitive estimation test. *Archives of Clinical Neuropsychology*, 19(24), 835-846.
- Burger, I. (2010). Verkeerd ingeschat?! De invloed van schattend rekenen, tijd en persoonlijke karakteristieken op kwantitatief schatten. Masterthese, Universiteit Utrecht.
- Della Sala, S., MacPherson, S.E., Philips, L.H., Sacco, L., & Spinnler, H. (2003). How many camels are there in Italy? Cognitive estimates standardised on the Italian population. *Neurological Science*, 24(10), 10-15.
- Dijksterhuis, A., Bos, M.W., Nordgren, L.F., & van Baaren, R.B. (2006). On Making the Right Choice: The Deliberation-Without-Attention Effect. *Science*, 311(5763), 1005-1007. [DOI:10.1126/science.1121629]
- Dijksterhuis, A., & Nordgren, L.F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science*, 2(1), 95-109.
- Gabrieli, J. , Fleischman, D., Keane, M., Reminger, S. & Morell, F. (1995). Double dissociation between memory systems underlying explicit and implicit memory in the human brain. *Psychological Science*, 6, 76-82.
- González-Vallejo, C. & Phillips, N. (2010). Predicting soccer matches: A reassessment of the benefit of unconscious thinking. *Judgment and Decision Making*, 5 (3), 200-206.
- Lassiter, G. D., Lindberg, M. J., Gonzalez-Vallejo, C., Belleza, F. S., & Phillips, N. D. (2009). The deliberation-without-attention effect: Evidence for an artifactual interpretation. *Psychological Science*, 20, 671–675.
- Lemaire, P., Arnaud, L., & Lecacheur, M. (2004). Adults' age-related differences in adaptivity of strategy choices: Evidence from computational estimation. *Psychology and*

*Aging*, 19(3), 467-481.

Nordgren, L.F., Bos, M.W., & Dijksterhuis, A. (2011). The best of both worlds: Integrating conscious and unconscious thought best solves complex decisions. *Journal of Experimental social psychology*, *Journal of experimental social psychology*, 47(2), 509-511

Paritosh, P.K. & Klenk, M.E. (2006). Cognitive Processes in Quantitative Estimation: Analogical Anchors and Causal Adjustment. To appear in the *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, Vancouver.

Rhymer, K.N., Skinner, C.H., Jackson, S., McNeill, S., Smith, T., & Jackson, B. (2002). The 1-minute explicit timing intervention: The influence of mathematics problem difficulty. *Journal of Instructional Psychology*, 29(4), 305-311.

Strick, M., Dijksterhuis, A., Bos, M.W., Sjoerdsma, A., Van Baaren, R.B., & Nordgren, L.F. (2010). A meta-analysis on unconscious thought effects. Manuscript in preparation. See [www.unconsciouslab.com](http://www.unconsciouslab.com).

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185, 1124-1131.

Zimbardo, P.G., Johnson, R.L., & McCann, V. (2009). *Psychologie een Inleiding*. Amsterdam: Pearson Education Benelux bv.