



Verkeerd ingeschat?!

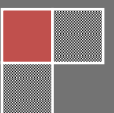
De invloed van schattend rekenen, tijd en persoonlijke karakteristieken op kwantitatief schatten.

Juni 2010

Auteur: Ingrid Burger
Studentnummer: 3332357

Beoordelaars: Casper Hulshof
Jeroen Janssen

Universiteit Utrecht



Samenvatting

Het maken van adequate schattingen is een belangrijke functie in het dagelijkse leven. Schatten is het verkrijgen van adequate antwoorden op vragen waarvoor geen specifiek antwoord beschikbaar is. Relevante en algemene kennis met betrekking tot het onderwerp is van groot belang om een dergelijke schatting te kunnen maken. Er zijn verschillende vormen van schatten. In dit onderzoek zijn twee vormen betrokken, rekenvaardig schatten en kwantitatief schatten. Er is onderzocht of kwantitatief schatten samenhangt met rekenvaardig schatten. Daarnaast is nagegaan welke invloed tijdslimiet en persoonlijke karakteristieken hebben op het maken van kwantitatieve schattingen. Met behulp van de online surveytool Limesurvey hebben 237 respondenten het meetinstrument ingevuld (met of zonder tijdslimiet). Uit de analyses blijkt een kleine 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 niet significant. Onderwijsniveau heeft als enige persoonlijke kenmerk invloed op rekenvaardig schatten. De resultaten geven eveneens aan dat verder onderzoek naar de strategieën van kwantitatief schatten en betrouwbaarheid van de gebruikte test nodig is.

§ 1 Inleiding op de probleemstelling

Weet u hoeveel pitten er in een watermeloen zitten? Om antwoord te geven op dergelijke vragen moet een schatting worden gemaakt. Schatten, of *cognitive estimation*, wordt omschreven als het verkrijgen van redelijke antwoorden op vragen waarvoor geen specifiek antwoord beschikbaar is (Barabassy, Beinhoff & Riepe, 2010; Bullard et al., 2004). Exacte kennis op deze vragen is niet direct toegankelijk. Het maken van adequate schattingen is een belangrijke functie in het dagelijkse leven (Barabassy et al., 2010; Della Sala, MacPherson, Philips, Sacco & Spinnler, 2003). Schatten is van belang om te kunnen inspelen op de omgeving (Forrester & Latham, 1990). Het uitvoeren van dagelijkse taken en het maken van verschillende beslissingen zijn gebaseerd op schattingen. Schatten is onder andere een onderdeel van het koken van een maaltijd, het oversteken van een weg, het autorijden, het reizen naar een afspraak of het uitrekenen van de kosten aan boodschappen. Het

uitrekenen van de exacte berekening is niet in elke situatie haalbaar. Factoren zoals de benodigde tijd of de complexiteit van de schattingssituatie hebben invloed op het maken van adequate schattingen.

Er zijn verschillende vormen van schatten die gekenmerkt worden door verschillende processen, strategieën en uitkomsten. In de wetenschappelijke literatuur worden de vormen rekenvaardig schatten, numeriek schatten en kwantitatief schatten nader onderzocht. Dit onderzoek is ingekaderd tot rekenvaardig schatten en kwantitatief schatten. Rekenvaardig schatten is het vinden van een geschikt antwoord op rekenkundige problemen zonder het antwoord exact uit te rekenen (Lemaire, Arnaud & Lecacheur, 2004). Bij kwantitatieve schattingen worden getallen toegeschreven aan objecten, gebeurtenissen of abstracties (Brown & Siegler, 1993). Om een kwantitatieve schatting te maken, gebruiken schatters domeinspecifieke kennis uit het geheugen (Brand, Kalbe, Fujiwara, Huber & Markowitsch, 2003; Bullard et al., 2004; Della Sala et al., 2003). Brown en Siegler (1993) menen dat er meerdere soorten kennis en strategieën vooraf gaan aan het maken van kwantitatieve schattingen, welke zij hebben vormgegeven in het Metrics and Mappings model. In dit model zijn naast domeinspecifieke kennis ook vuistregels en numerieke inductie opgenomen als strategieën, die leiden tot een kwantitatieve schatting. Het model van Siegler en Brown ziet rekenkundige bewerkingen niet als een strategie om tot een schatting te komen. De vraag is echter of het maken van kwantitatieve schattingen alleen kennis binnen het betreffende domein vraagt. Het uitvoeren van berekeningen kan eveneens een mogelijke strategie zijn om tot een adequaat schatting te komen. In dit onderzoek is nagegaan in hoeverre respondenten daadwerkelijk een antwoord exact proberen uit te rekenen. Daarnaast is met dit onderzoek de samenhang tussen rekenvaardig schatten en kwantitatief schatten nader onderzocht, het betreft hier de eerste deelvraag.

Er zijn situaties waarbij men weinig tijd heeft om tot een schatting te komen aangezien men snel moet reageren op wat in de omgeving plaatsvindt. Volgens het onderzoek van Rhymer et al. (2002) heeft een tijdslimiet geen invloed op het correct uitrekenen van sommen. De invloed van tijd op het maken van kwantitatief schatten is echter nog niet nader onderzocht. Een tijdslimiet heeft mogelijk een effect op het maken van kwantitatieve schattingen. Dit is het tweede deelaspect wat in dit onderzoek nader is onderzocht. Wanneer er daadwerkelijk sprake is van een effect, sluit dit aan bij de

bevindingen van Dijksterhuis, Bos, Nordgren en Van Baaren (2006), dat complexe situaties om beslissingen vraagt met onbewuste overwegingen.

Uit onderzoeken blijkt dat persoonlijke karakteristieken zoals geslacht, leeftijd of onderwijsniveau, het maken van kwantitatieve schattingen lijken te beïnvloeden (Della Sala et al., 2003; Spencer & Johnson-Greene, 2009). Bij dit onderzoek is nagegaan of persoonlijke karakteristieken invloed hebben op het maken van kwantitatieve schattingen (deelvraag 3).

Uit deze overwegingen is de volgende onderzoeksvraag geformuleerd welke centraal staat bij dit onderzoek: *hangt het kwantitatief schatten samen met rekenvaardig schatten en in welke mate beïnvloeden de factoren tijdslimiet en persoonlijke karakteristieken de schattingsvaardigheden?* Rekenvaardig schatten is getest met behulp van tien rekensommen uit twee rekenmethoden uit groep acht van het basisonderwijs. Kwantitatief schatten is met behulp van een vertaling van de Biber Cognitive Estimation Test [BCET] (Bullard et al., 2004) nagegaan.

Het is niet de doelstelling van dit onderzoek de betrouwbaarheid en validiteit van de test [BCET] te toetsen. Silverberg, Hanks en McKay (2007) vinden de BCET een verbetering ten opzichte van andere kwantitatieve schattingstesten. Andere testen bevatten items waar gevraagd wordt naar feitenkennis of waarvan het antwoord exact te berekenen is. Daarnaast is de BCET bij verschillende onderzoeken als meetinstrument gehanteerd. Ten slotte vraagt het testen en eventueel verbeteren van de BCET een ander soort onderzoek welke niet te gerelateerd kon worden met dit onderzoek.

Onderzoeken naar kwantitatief schatten worden over het algemeen binnen de neuropsychologische context verricht bijvoorbeeld bij Alzheimer patiënten (Della Sala et al., 2004), Parkinson patiënten (Bullard et al., 2004) of jongeren met een stoornis binnen het autistische spectrum (Liss, Fein, Bullard & Robins, 2000). Dit onderzoek is vernieuwend omdat 'gezonde' respondenten zijn betrokken bij dit onderzoek. In de huidige onderzoeken wordt tijdslimiet niet als factor op het maken van kwantitatieve schattingen betrokken. Dit onderzoek heeft tijdslimiet als mogelijke beïnvloedende factor nader onderzocht. Tijdslimiet kan eveneens een nieuw criterium vormen voor de uitvoering van kwantitatieve schattingstesten. Bevindingen uit dit onderzoek kunnen een belangrijke bijdrage leveren aan wetenschappelijk onderzoek wanneer blijkt dat rekenvaardig schatten samenhangt met kwantitatief schatten.

Dit artikel gaat verder met de theoretische achtergrond van schatten. Eerst worden verschillende vormen van schatten en het proces van kwantitatief schatten nader toegelicht. Het theoretisch kader wordt afgesloten met een beschrijving van de factoren die in dit onderzoek zijn onderzocht. Vervolgens wordt de methode van het onderzoek beschreven. Hierna worden de resultaten met betrekking tot de drie deelvragen weergegeven. Met behulp van deze resultaten wordt in de conclusie antwoord gegeven op de onderzoeksvraag. Dit artikel sluit af met discussiepunten en aanbevelingen met betrekking tot het huidige kennisbestand en methodologie van dit onderzoek.

§ 2 Theoretische achtergrond

§ 2.1 Vormen van schatten

Schattingssituaties hebben allen gemeen dat een adequaat antwoord vaak in een korte tijd gevonden moet worden (Rubenstein, 1986). Tevens kunnen schattingssituaties op diverse vlakken verschillen. Zo vraagt de ene situatie meten, terwijl de andere situatie vraagt om uitrekenen, onderzoeken of voorspellen. Verschillen zijn er eveneens met betrekking tot de complexiteit, de context, de hoeveelheid informatie die beschikbaar is en de getallen die betrokken zijn om een adequate schatting te kunnen maken. Ten slotte verschillen schattingssituaties tot het type antwoord dat gewenst is. De volgende drie vormen van schatten worden in de wetenschappelijke literatuur beschreven en worden nader toegelicht. Het gaat om *computational estimation*, *numerical estimation* en *quantitative estimation*.

De eerste vorm van schatten is *computational estimation*, ‘rekenvaardig schatten’. Rekenvaardig schatten is het vinden van een geschikt antwoord op rekenkundige problemen zonder het exacte antwoord uit te rekenen (Lemaire et al., 2004). Rekenvaardig schatten sluit aan bij het schattend rekenen of het handig rekenen van de rekenmethoden in het huidige basisonderwijs, met sommen zoals $47 \times 53 \approx 2500$ en $146 + 69 + 48 \approx 260$. Tijdens het schatten worden verschillende strategieën gebruikt, zoals de afrondstrategie waarbij getallen naar boven of onder worden afgerond (Dowker, 1992 zoals geciteerd in Klunder, 2005). Een andere strategie is de clusterstrategie, waarbij getallen met ongeveer dezelfde waarden worden samengenomen. Het snel en nauwkeurig kunnen

uitvoeren van rekenvaardige schattingen kent twee voordelen (Star & Rittle, 2009). Ten eerste kan de schatter het antwoord controleren op redelijkheid, wanneer gebruik wordt gemaakt van meerdere strategieën. Ten tweede helpt rekenvaardig schatten tot een beter begrip te komen van de plaatswaarde van getallen, rekenkundige operaties en getalbegrip (Beishuizen, Van Putten & Van Mulken, 1997).

Numeriek schatten (*numerical estimation*) is de tweede vorm van schatten die benoemd wordt in de wetenschappelijke literatuur. Siegler en Opfer (2003) hebben bijvoorbeeld onderzoek verricht naar numeriek schatten. Bij dit onderzoek moest een getal op een lege getallenlijn (tot 100 of tot 1000) worden geplaatst of werd geschat welk getal bij een streep op de getallenlijn hoorde. Volgens Siegler en Opfer (2003) was deze design voor numeriek schatten geschikt voor de respondenten (jonge kinderen) aangezien kennis van meeteenheden bij deze design niet van belang is. Daarnaast zijn kinderen bekend met getallenlijnen en is het mogelijk een grotere range aan getallen te gebruiken. Onderzoeken naar numeriek schatten hebben gemeen dat bij het maken van deze schattingen gebruik wordt gemaakt van een getallenlijn of de verhouding ten opzichte van een ander getal. Het verschil met rekenvaardig schatten is dat het uitvoeren van een berekening bij numeriek schatten niet altijd noodzakelijk is.

Als laatste vorm wordt *quantitative estimation*, ook wel kwantitatief schatten, onderzocht en beschreven. Deze vorm van schatten wordt omschreven als het proces waarbij mensen een numerieke waarde toeschrijven aan objecten, gebeurtenissen of abstracties (Brown & Siegler, 1993). Deze numerieke waarden worden door middel van relevante en algemene kennis gegeven (Bullard et al., 2004; Della Sala et al., 2003). Kwantitatief schatten verschilt ten opzichte van rekenvaardig schatten met betrekking tot het toepassen van berekeningen. Kwantitatief schatten richt zich op het toepassen van domeinkennis en rekenvaardig schatten op het berekenen van het antwoord. Het verschil met numeriek schatten is dat kwantitatieve schattingen meerdere domeinen betreft zoals afstand, tijd en gewicht. Bij numerieke schattingen gaat het om schattingen met behulp van een getallenlijn en is hierdoor abstracter dan kwantitatief schatten. Bij dit onderzoek gaat het hoofdzakelijk om kwantitatief schatten. Rekenvaardig schatten wordt als mogelijk factor betrokken en staat hiermee minder centraal in dit onderzoek. Het proces om tot kwantitatieve schattingen te komen wordt in de volgende paragraaf nader uitgelegd.

§2.2 *Proces van kwantitatief schatten*

Om tot een schatting te komen gaat een opeenvolgend proces vooraf. Als eerst analyseert men het probleem (Liss et al., 2000). Daarna identificeert en activeert men de relevante kennisbasis die betrekking heeft op de schattings situatie (Bullard et al., 2004). Uit deze relevante kennisbasis haalt men specifieke feiten of eerder uitgevoerde schattingen naar boven (Bullard et al., 2004; Liss et al., 2000). Vervolgens bedenkt men een geschikte manipulatie voor deze feiten of schattingen en voert men deze manipulatie uit. Deze manipulaties noemen Shallice en Evans (1978) cognitieve plannen. Echter, wat deze manipulaties daadwerkelijk inhouden beschrijven Shallice en Evans niet. Ten slotte worden de resultaten uit deze manipulaties vergeleken met de kennis van de wereld. Met deze vergelijking is het mogelijk te bepalen of deze schattingen redelijk en realistisch zijn. Het resultaat, in dit geval het antwoord op de schattings situatie, wordt getoetst op de realiteit (Shallice & Evans, 1978). De uitgevoerde cognitieve manipulatie wordt eveneens gecontroleerd omdat de manipulatie geschikt of ongeschikt kan zijn. Dit volledige proces kan herhaald worden wanneer men tot een beter antwoord wenst te komen (Bullard et al., 2004). Het herhalen van dit proces noemen Liss et al. (2000) zelfcorrectie. Een goede schatting is in de meeste contexten een '*best-fit*' antwoord waarbij rekening wordt gehouden met de essentiële achtergrondfactoren van een bepaalde context (Forrester & Latham, 1990).

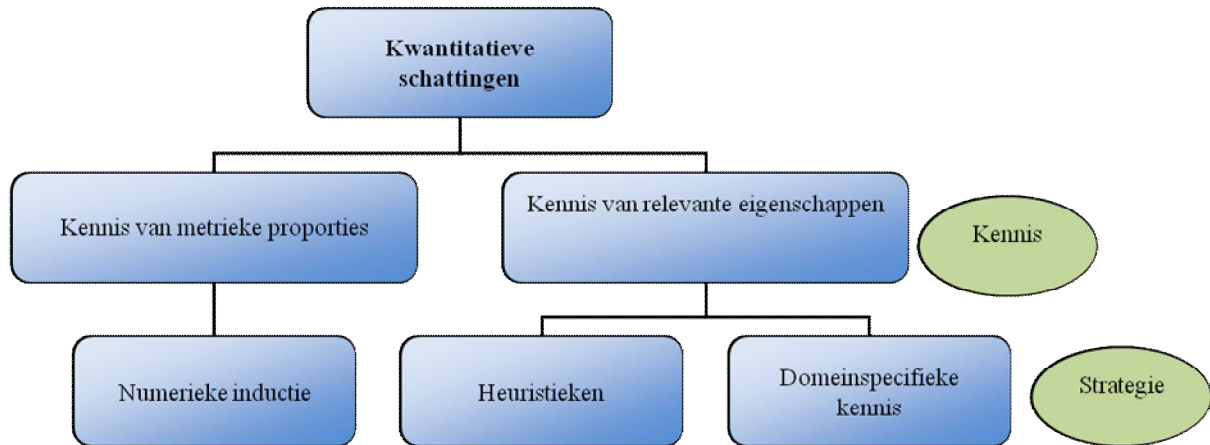
Brown en Siegler (1993) hebben de kennis om te komen tot een schatting uitgebreider beschreven. Volgens hen zijn er twee soorten kennis met verschillende strategieën die gebruikt worden bij het maken van kwantitatieve schattingen. Dit hebben Brown en Siegler vormgegeven in het Metrics and Mappings model.

§ 2.2.1 *Metrics and Mappings model*

Brown en Siegler (1993) zijn van mening dat schattingen gemaakt worden door middel van kennis van metrieke propriëties (*metric propriëties*) en kennis van relevante eigenschappen (*mappings propriëties*). Kennis van metrieke propriëties is de statistische of numerieke kennis van de schattings situatie, zoals het gemiddelde of de mediaan. Bij de kennis van relevante eigenschappen gaat het om de ordinale relaties tussen verschillende objecten in een domein, waarbij het ene object een hogere waarde krijgt

dan het andere object (Von Helversen & Rieskamp, 2008). Het model van Siegler en Brown (1993) is weergegeven in figuur één.

Figuur 1. Metrics and Mappings model¹



¹Overgenomen uit "Metrics and mappings: A framework for understanding real-world quantitative estimation," door N.R. Brown, en R.S. Siegler, 1993, *Psychological Review*, 100(3), p. 514. Copyright 1993 door American Psychological Association Inc.

Om tot deze twee soorten kennis te komen, zijn er drie strategieën van belang namelijk het toepassen van numerieke inductie, heuristieken of domeinspecifieke kennis. Bij numerieke inductie gaat het om het vaststellen van een range: wat is het antwoord in ieder geval niet en tussen welke twee uitersten ligt het antwoord mogelijk wel? Op basis van deze range gebruikt de schatter het gemiddelde of de mediaan om een antwoord te geven op de schattingsituatie. Wanneer men schat hoe vaak men zijn tanden kan poetsen met een tube tandpasta, kunnen schatters een range aangeven van 20 – 50. Buiten deze range ligt het antwoord volgens hen in ieder geval niet. Bij dit proces heeft de schatter kennis nodig van het numerieke stelsel en de meeteenheden. Praktisch gezien is het van belang dat de schatter weet dat 1000 groter is dan 500 en één kilo zwaarder is dan 100 gram.

Het toepassen van heuristieken is de tweede strategie. Het gaat hier om het toepassen van vuistregels waar de schatting op gebaseerd wordt. Murray en Brown (2009) hebben een onderzoek verricht naar het schatten van autoprijzen. Bij dit onderzoek werden vuistregels zoals automerk en autovariant (cabrio, sport en gezinsauto) betrokken bij het schatten van autoprijzen. Het nadeel van het schatten met behulp van deze strategie is dat heuristieken de schatting onbedoeld in een bepaalde

richting kunnen sturen (Brown & Siegler, 1993). In het onderzoek van Murray en Brown (2009) werd een Honda cabrio ten onrechte goedkoper geschat dan een BMW cabrio. De vuistregel die respondenten hadden toegepast is dat Honda goedkoper is dan BMW, ongeacht het autotype.

De laatste mogelijke strategie is het toepassen van domeinspecifieke kennis. Het gaat bij deze benadering om het ophalen en maken van beslissingen met specifieke en relevante aspecten van de kennis binnen het domein. Er wordt bij deze strategie geen gebruik gemaakt van vuistregels. Wanneer men schat hoeveel twaalf normale appels wegen, zal men de ervaring gebruiken van het wegen en kopen van appels in de supermarkt. Bij het schatten van de lengte van een spin, kan de schatter de grootte van een spin visualiseren (Shallice & Evans, 1978).

Brown en Siegler (1993) zien deze drie strategieën niet als geïsoleerde benaderingen die het schattingsproces beïnvloeden. Domeinspecifieke kennis kan bijvoorbeeld de numerieke inductie ondersteunen. Een kilo gehakt is meer dan 500 gram en deze relatie kan ook gelegd worden naar kilometers en meters. Daarom hebben Brown en Siegler een model ontwikkeld waarbij de drie benaderingen geïntegreerd zijn.

Dat verschillende soorten kennis en strategieën van belang zijn bij het schatten hebben Von Helversen en Rieskamp (2008) toegelicht met een praktisch voorbeeld. Het gaat bij deze toelichting om het schatten van verkoopprijzen van twee nieuwe mobiele telefoons. Het schatten van prijzen gaat aan de hand van de kennis van eigenschappen (*mappings*) zoals gewicht, soort beeldscherm, aanwezige digitale camera en toegang tot internet. De eigenschappen van de twee nieuwe mobiele telefoons worden vergeleken met de eigenschappen van mobiele telefoons die in het verleden zijn gekocht. De prijzen van de oude mobiele telefoons worden vergeleken met de eigenschappen van de nieuwe telefoons. De verhouding tussen prijs en eigenschappen van de oude mobiele telefoons worden toegepast op de twee nieuwe telefoons (*metrics*). Op basis van deze vergelijking komt men tot een schatting.

Het Metrics and Mappings model van Brown en Siegler (1993) beschrijft drie soorten strategieën om tot een kwantitatieve schatting te komen. De vraag is echter of het uitvoeren van rekenkundige bewerkingen ook als een strategie gezien kan worden. Op de vraag hoeveel een paard en wagen in één uur kan rijden, zal iemand kunnen berekenen hoe hard iemand fietst en dit kunnen

doorberekenen naar hoe hard een paard en wagen zal gaan. Dit kan betekenen dat rekenkundige bewerkingen als strategie voor kwantitatieve schattingen een toevoeging voor het model kan zijn. Voordat rekenvaardig schatten als mogelijke strategie verder wordt onderzocht, is het belangrijk eerst na te gaan of rekenvaardig schatten samenhangt met kwantitatief schatten. Siegler en Booth (2004) hebben al aangetoond dat de score op numeriek schatten samenhangt met de score op rekenkundige berekeningen. In dit onderzoek is nagegaan of kwantitatief schatten samenhangt met rekenvaardig schatten en is hiermee de eerste deelvraag.

§ 2.3 Beïnvloedende factoren bij kwantitatief schatten

Volgens Shallice en Evans (1978) is het uitvoeren van schattingen een cognitieve activiteit welke wordt aangestuurd door de frontale kwab. Wanneer men niet in staat is om te schatten, wordt dit onvermogen om te schatten gekoppeld aan het disfunctioneren van de frontale kwab (Barabassy et al., 2010). Mensen met schade aan de frontale kwab hebben moeite om een cognitief plan te formeren en in te zetten of tot een strategie te komen om een probleem op te lossen (Smith & Milner, 1984). Bij schatten zijn verschillende complexe cognitieve functies van belang zoals het activeren van het geheugen, ophalen van kennis uit het geheugen, planning, mentale controle, zelfmonitoring en zelfcorrectie (Bullard et al., 2004). Volgens Bullard et al. (2004) zijn deze functies beperkt wanneer er schade is ontstaan in de frontale kwab. Testen naar kwantitatief schatten worden binnen de neuropsychologie uitgevoerd. In tegenstelling tot bestaande onderzoeken is in dit onderzoek nagegaan of verschillende factoren bij gezonde respondenten van invloed zijn op het maken van adequate schattingen. Het gaat hierbij om de factoren persoonlijke karakteristieken en tijdslimiet.

§ 2.3.1 Tijd als mogelijke factor kwantitatief schatten

Volgens Dijksterhuis et al. (2006) is niet elke situatie geschikt om met bewuste overwegingen tot een keuze te komen. Dijksterhuis et al. (2006) menen dat in bepaalde situaties het zelfs beter is om keuzes met onbewuste overwegingen te maken. *Onconscious thought*, wordt door hen omschreven als de gedachten of de overwegingen zonder bewuste aandacht te geven aan het probleem. Bij eenvoudige situaties is het goed om bewuste overwegingen te maken. Het tegenovergestelde geldt voor complexe

situaties waarbij het beter is om een keuze te maken met onbewuste overwegingen. Bewuste overwegingen hebben namelijk een beperkte capaciteit om informatie op te nemen en zijn hierdoor niet geschikt voor complexe situaties.

Het maken van schattingen met bewuste of onbewuste overwegingen is nog niet onderzocht binnen de context van kwantitatief schatten. Dat neemt niet weg dat bewuste overwegingen dan wel onbewuste overwegingen een rol kunnen spelen. Tijdslimieten (beperkte denktijd) kunnen het maken van schattingen met onbewuste overwegingen mogelijk afdwingen. Wanneer iemand alle tijd krijgt om tot een antwoord te komen, kunnen alle mogelijke overwegingen uitgevoerd worden. Indien deze tijd niet gegeven wordt, is het niet mogelijk deze bewuste overwegingen te maken. In deze situatie kan men een antwoord geven op basis van onbewuste overwegingen Dijksterhuis et al. zeggen echter het volgende: “Men kan niet rekenen zonder bewuste aandacht” (Dijksterhuis et al, 2006, p. 1006). Dit sluit aan bij het regel principe, waarmee bedoeld wordt dat bewuste overwegingen strikte regels volgen en nauwkeuriger zijn dan onbewuste overwegingen (Dijksterhuis & Nordgren, 2006). Onbewuste overwegingen geven volgens Dijksterhuis en Nordgren (2006) echter ruwe schattingen.

Bij dit onderzoek is tijd als beïnvloedende factor betrokken bij rekenvaardig schatten en kwantitatief schatten. Het hanteren van tijdslimieten bij het uitvoeren van rekenkundige berekeningen hebben volgens Rhymer et al. (2002) over het algemeen een positieve invloed. Zij hebben onderzoek verricht onder leerlingen van groep acht waarbij leerlingen sommen met een tijdslimiet en zonder een tijdslimiet moesten maken. Uit de analyse kwam naar voren dat de leerlingen uit de experimentele conditie, rekenen met een tijdslimiet van één minuut, significant meer sommen per minuut konden oplossen dan de leerlingen zonder een tijdslimiet. Het percentage correcte antwoorden was alleen significant hoger voor de conditie met tijdslimiet op de bewerkingen optellen en aftrekken. Andere bewerkingen zoals vermenigvuldigen of delen verschilden niet significant van elkaar.

De tweede deelvraag richt zich op de invloed van een tijdslimiet op rekenvaardig schatten en kwantitatief schatten. Indien dit effect te leggen is, kan dat duiden op de bewuste of onbewuste overwegingen.

§ 2.3.2 Persoonlijke karakteristieken als factor kwantitatief schatten

Persoonlijke karakteristieken zoals geslacht, leeftijd of onderwijsniveau lijken het maken van kwantitatieve schattingen te beïnvloeden. Vrouwen scoorden significant lager dan de mannen op de kwantitatieve schattingstest van Della Sala et al. (2003) en Spencer en Johnson-Greene (2009). In het onderzoek van Gillepsie, Evans, Gardener en Bowen (2002) waren er echter geen significante verschillen tussen mannen en vrouwen.

Uitkomsten op de schattingstest kunnen correleren met leeftijd. De correlatie was echter vrij laag bij het onderzoek van Della Sala et al. (2003). In het onderzoek van Liss et al. (2000) voorspelde leeftijd binnen de gezonde controle groep voor een groot deel de score op de schattingstest, namelijk 42%. Volgens Della Sala et al. (2004) zijn er geen significante verschillen tussen jonge volwassenen en ouderen. Onderzoeken van Gillepsie et al. (2002) en Spencer en Johnson-Greene (2009) beaamen eveneens dat er geen significante relatie is tussen leeftijd en de uitkomst op de schattingstest.

Wanneer onderwijsniveau is meegenomen in het onderzoek, bleek dat het onderwijsniveau correleerde met de uitkomsten op de schattingstest bij zowel het onderzoek van Della Sala et al. (2003) als Spencer en Johnson-Greene (2009).

In dit onderzoek is nagegaan of in deze context de persoonlijke karakteristieken een beïnvloedende factor zijn op het maken van kwantitatieve schattingen en betreft hiermee de derde en laatste deelvraag.

§ 3 Methode van onderzoek

§ 3.1 Respondenten

In totaal hebben 239 respondenten het volledige meetinstrument doorlopen. Twee respondenten zijn uit het bestand gehaald. Een respondent had de leeftijd van vijftien jaar welke onder de grens lag van het minimum van zestien jaar. De andere respondent had bij het gedeelte van kwantitatief schatten commentaar op de vragen gegeven in plaats van numerieke antwoorden, zoals “dit vind ik een onzin vraag” of “dat ligt aan heel veel dingen”.

In totaal zijn er 237 volledige testen betrokken bij de analyses. Het betreft 146 vrouwen en 91 mannen met een gemiddelde leeftijd van 31,42 jaar ($SD = 12,50$). De verdeling van het aantal respondenten over de twee groepen is vrijwel gelijk, 120 respondenten bij de tijdslijmiegroep en 117 respondenten bij de groep zonder tijd. Gegevens van de respondenten per groep zijn weergegeven in tabel één.

Naast 239 volledig doorlopen testen, zijn er in totaal 153 afgebroken testen. Bij de tijdslijmiegroep zijn dit 100 en de andere 53 testen bij de groep zonder tijdslijmie. Deze testen zijn niet meegenomen bij de analyses.

Tabel 1

Gegevens respondenten per groep

Persoonlijke gegevens		Groep tijdslijmie	Groep zonder tijd	Totaal
Geslacht	mannelijk	44	47	91
	vrouwelijk	76	70	146
Leeftijd	gemiddeld	32,69 ($SD = 13,40$)	30,11 ($SD = 11,41$)	31,42 ($SD = 12,50$)
Onderwijsniveau	lbo	3	0	3
	mavo/vmbo	9	4	13
	havo	17	19	36
	vwo	12	10	22
	mbo	25	24	49
	hbo	40	45	85
	wo	14	15	29

§ 3.2 *Instrument*

Het gebruikte instrument dat bij dit onderzoek gebruikt is, bestaat uit drie delen. De dataverzameling is gestart met het verzamelen van informatie over de respondenten. Hierbij werd gevraagd naar geslacht, leeftijd en hoogst genoten opleiding.

Het meetinstrument vervolgde met tien sommen binnen het domein van rekenvaardig schatten. Deze sommen zijn geselecteerd uit twee rekenmethoden voor groep acht van het basisonderwijs. De rekenmethoden waren *Wereld in Getallen* (Huitema, Van der Klis, Timmermans & Erich, 2001) en *Pluspunt* (Vijgen & Van Broekhoven, 2000). Er is gestreefd naar een goede verdeling tussen de basisuitvoeringen: optellen, aftrekken, vermenigvuldigen en delen. Daarbij zijn verschillende grootheden gebruikt, zoals tientallen, honderdtallen, duizendtallen en getallen met decimalen. De sommen en de antwoorden welke als correct worden beschouwd, zijn weergegeven in tabel twee. De range is gebaseerd op de afrondstrategie op beide getallen van Dowker (1992, zoals geciteerd in Klunder, 2005).

Tabel 2

Onderdeel rekenvaardig schatten dataverzameling

Som	Exacte respons	Afrondstrategie	Range correcte respons
797 - 249	548	800 - 250 = 550	545 - 550
5,15 x 12,93	66,59	5 x 13 = 65	65 - 67
418 : 8	52,25	420: 8 \approx 52 (= 52,50)	50 - 54
3487 + 2829	6316	3500 + 3000 = 6500	6300 - 6500
87,15 - 19,4 - 0,88	66,87	87 - 20 - 1 = 66	65 - 67
51 x 12,992	662,59	50 x 13 = 650	650 - 665
4835 + 975 + 1214	7024	5000 + 1000 + 1200 = 7200	7000 - 7200
6,18 + 12,57 + 19,38	38,13	6 + 13 + 20 = 39	37 - 40
48 x 72	3456	50 x 70 = 3500	3400 - 3500
8717 : 23	379	8700: 20 \approx 400 (= 435)	370 - 400

Na het gedeelte van rekenvaardig schatten, zijn de respondenten gevraagd in hoeverre zij de antwoorden op de sommen exact hadden berekend of geschat. Deze vragen dienen als een soort controle en reflectie op het antwoordproces. In het vervolg van dit onderzoek zal met 'aanpak van respondenten' worden verwezen naar deze twee vragen.

Het laatste gedeelte van de test betrof het gedeelte van kwantitatief schatten. Hiervoor is gebruik gemaakt van een vertaling van de *Biber Cognitive Estimation Test* [BCET]) van Bullard et al. (2004). Dit is één van de testen die gebruikt wordt binnen wetenschappelijk onderzoek en vooral op het gebied van de neuropsychologie. De test bestaat uit twintig items waarbij vier categorieën worden bevraagd: tijd, hoeveelheid, gewicht en afstand. Volgens Silverberg et al. (2007) is de schattingstest van Bullard et al. (2004) een verbetering ten opzichte van de tekortkomingen van eerdere onderzoeken met schattingstesten. Volgens hen beschikt de BCET over betere normen en bevat de test items waar een echt antwoord niet beschikbaar is. Daarnaast is het volgens Silverberg et al. een verbetering aangezien de test de meeteenheden (kilogrammen, meters et cetera.) niet aanbiedt, zodat de schatter gestuurd wordt in zijn schatting en is er sprake van een gelijke verdeling van de verschillende domeinen. Het meetinstrument van Della Sala et al. (2003) is niet gekozen omdat deze test een aantal items bevat die met feitenkennis of een exacte berekening te beantwoorden zijn. “Hoeveel weken zitten er in een jaar?” of “wat is de oppervlakte van een tweepersoons matrashoes?” (p. 12) zijn hiervan voorbeelden. Bij het onderzoek Gillepsie et al. (2010) en Brand et al. (2003) moesten respondenten schattingen maken waarbij zij werden ondersteund met figuren. “Hoe hoog is deze voordeur?”, “hoeveel snoepjes zitten er in deze verpakkingen?” (Gillepsie et al., 2010, p. 293) of “hoe lang is deze vlieg” (Brand et al., 2003, p. 577) zijn voorbeelden van gehanteerde items. Het is af te vragen of deze vorm van schatten domeinafhankelijke kennis vraagt van de respondenten. Daarnaast is de repliceerbaarheid van dit soort onderzoeken lastig vanwege het achterwege blijven van de figuren. De BCET is daarentegen bij verschillende onderzoeken als meetinstrument gebruikt, zoals bij het onderzoek van Liss et al. (2000) en Silverberg et al. (2007). Bij beide onderzoeken, binnen een neuropsychologische context, had de BCET een matige betrouwbaarheid (Cronbach’s alpha .62 en .61). Het gaat bij dit onderzoek echter om een nieuwe context en andere onderzoeksgroepen. Er kan vanuit worden gegaan dat de BCET geschikt is om te gebruiken bij onderzoeken naar kwantitatieve schattingen in nieuwe contexten. Silverberg et al. (2007) vinden het zelfs nadelig dat bij onderzoek naar kwantitatief schatten telkens nieuwe testen worden ontwikkeld en andere scoringsmethoden worden gehanteerd. Volgens Silverberg et al. hebben deze aanpassingen een slechte invloed op de

validiteit van de schattingstesten. Het gevolg van verschillende testen en diverse aanpassingen hebben geleid tot een lage correlatie tussen diverse schattingstesten (Gillepsie et al., 2002).

De BCET is voor dit onderzoek vertaald naar het Nederlands (tabel drie). Bij de vertaling is een Amerikaanse context veranderd in een Nederlandse context. Vraag 19 is vervangen door een vraag binnen dezelfde categorie uit het onderzoek van Della Sala et al. (2003). De oorspronkelijke vraag is een vraag welke gevoelig kan liggen of onbedoelde reacties bij de respondenten kan oproepen (hoe zwaar is de zwaarste man van de Verenigde Staten/Nederland). De instructies welke Bullard et al. (2004) hebben gebruikt, zijn in dit onderzoek eveneens gehanteerd. Gelijk aan het gedeelte van rekenvaardig schatten is er gevraagd in hoeverre de respondent het antwoord heeft geprobeerd te berekenen of heeft geschat. Ook deze vragen zullen als ‘aanpak van respondenten’ worden bestempeld. Het volledige meetinstrument is opgenomen in Appendix A.

Het volledige meetinstrument is tijdens een pilot getest bij een aantal respondenten. Bij deze pilot zijn negen respondenten betrokken met een gemiddelde leeftijd van 25,57 jaar ($SD = 2,37$). Met deze pilot zijn de items op duidelijkheid en leesbaarheid getest. Daarnaast is het tijdslimiet bepaald. De pilot is schriftelijk afgenomen waarbij de tijd door de onderzoeker is gemeten. Voor het rekenvaardig schatten hebben de respondenten één minuut de tijd gekregen om antwoord te geven en vier minuten bij het kwantitatief schatten. Uit deze pilot is naar voren gekomen dat het item over zure melk specifieker kan worden weergegeven. De overige items zijn duidelijk en leesbaar. Het tijdslimiet van één minuut voor rekenvaardig schatten bleek te kort te zijn. Het tijdslimiet voor kwantitatief schatten was te ruim, alle respondenten waren ongeveer een minuut eerder klaar. In het definitieve design hebben de respondenten een tijdslimiet per vraag gekregen en niet per onderdeel. De bevindingen uit deze pilot zijn hierbij doorberekend.

Tabel 3

The Biber Cognitive Estimation Test

Categorie	Items in het Engels*	Item in het Nederlands
Hoeveelheid	1. How many seeds are there in a watermelon?	1. Hoeveel pitten zitten er in een watermelon?

Gewicht	2. How much does a telephone weigh?	2. Hoeveel weegt een mobiele telefoon?
Hoeveelheid	3. How many sticks of spaghetti are there in a one-pound package?	3. Hoeveel stokjes spaghetti zitten er in een pak van één pond?
Afstand	4. What is the distance an adult can walk in an afternoon?	4. Wat is de afstand dat een volwassene in een avond kan lopen?
Afstand	5. How high off a trampoline can a person jump?	5. Hoe hoog kan iemand op een trampoline springen?
Tijd	6. How long does it take a builder to construct an average-sized house?	6. Hoe lang duurt het bouwen van een woonhuis?
Gewicht	7. How much do a dozen, medium-sized apples weigh?	7. Hoeveel wegen 12 normale appels bij elkaar?
Afstand	8. How far could a horse pull a farm cart in one hour?	8. Wat is de afstand dat een paard en wagen in één uur kan afleggen?
Hoeveelheid	9. How many brushings can someone get from a large tube of toothpaste?	9. Hoe vaak kan je je tanden poetsen van een tube tandpasta?
Hoeveelheid	10. How many potato chips are there in a small, one-ounce bag?	10. Hoeveel chips zitten er in een gewone zak van 200 gram?
Tijd	11. How long would it take an adult to hand write a one-page letter?	11. Hoelang duurt het voor een volwassene om één kant van een A4 vol te schrijven?
Tijd	12. What is the age of the oldest living person in the United States?	12. Wat is de leeftijd van de oudste, levende persoon in Nederland?
Afstand	13. How long is a tablespoon?	13. Hoe lang is een eetlepel?
Gewicht	14. How much does a bridge (folding) chair weigh?	14. Hoeveel weegt een inklapbare tuinstoel?
Tijd	15. How long does it take to iron a shirt?	15. Hoeveel tijd kost het om een T-shirt te strijken?
Afstand	16. How long is a giraffe's neck?	16. Hoe lang is de nek van een giraffe?
Hoeveelheid	17. How many slices of bread are there in a one-pound loaf?	17. Hoeveel plakken brood zitten er in een brood van 500 gram?
Gewicht	18. How much does a pair of men's shoes weigh?	18. Hoeveel weegt een paar mannenschoenen?

Gewicht	19. How much does the fattest man in the United States weigh?	19. Hoe zwaar is een paard? ^a
Tijd	20. How long does it take for fresh milk to go sour in the refrigerator?	20. Hoe lang duurt het voordat verse melk na de houdbaarheidsdatum in de koelkast zuur wordt?

Noot. Overgenomen uit "The Biber Cognitive Estimation Test", S.E. Bullard, D. Fein, M.K. Gleeson, N. Tischer, R.L. Mapou, en E. Kaplan, 2004, *Archives of Clinical Neuropsychology*, 19(24), 835-846.

^a Overgenomen uit "How many camels are there in Italy? Cognitive estimates standardised on the Italian population", S. Della Sala, S.E., MacPherson, L.H. Philips, L. Sacco en H. Spinnler, 2003, *Neurological Science*, 24(10), 10-15.

§ 3.3 Design en procedure

De dataverzameling omvat twee groepen. De eerst groep betreft de uitvoering zonder tijdslimieten. Het gedeelte van rekenvaardig schatten en het kwantitatief schatten zijn door de respondenten op eigen tempo doorlopen. Het werd hierdoor mogelijk schattingen te maken met bewuste overwegingen. De tweede groep, betrof de dataverzameling met tijdslimieten. Respondenten binnen deze groep hebben een bepaalde tijd gekregen om zowel het rekenvaardig schatten, als het kwantitatief schatten uit te voeren. Bij deze groep moesten schattingen gemaakt worden met onbewuste overwegingen. Bij zowel het gedeelte van rekenvaardig schatten als kwantitatief schatten hebben de respondenten vijftien seconden per som of per vraag gekregen. De vragen met betrekking tot persoonlijke karakteristieken of de aanpak van de respondent werden zonder tijdslimiet doorlopen.

Het meetinstrument is op een internetsite (Limesurvey, 2010) ontwikkeld en geactiveerd. De respondenten hebben de link via de mail toegestuurd gekregen. Hierbij is gevraagd of de respondenten de test wilden doorsturen. Na anderhalve week kregen de respondenten is een herinnering toegestuurd. Wanneer de dataverzameling werd afgerond zijn deze gegevens overgenomen, verwerkt en geanalyseerd in SPSS, versie 17.

§ 3.4 Scoringsmethode

Nadat de data ingevoerd was, is voor de schattingstest de scoringsmethode van Axelrod en Millis (1994, zoals gebruikt bij het onderzoek van Silverberg et al., 2007) gehanteerd. Met een scoringsmethode wordt de manier bedoeld waarmee een antwoord als correct of als oncorrect wordt beoordeeld. Bij de scoringsmethode van Axelrod en Millis krijgen antwoorden onder het 2^{de} percentiel

en boven het 98^e percentiel de hoogste score, twee punten; antwoorden tussen 2^e - 15^e en 84^e - 98^e percentiel krijgen één punt en de antwoorden tussen 16^e en 84^e percentiel krijgen geen punten. Voor dit onderzoek werd de puntentoekenning omgewisseld om overeen te komen met de scoringsmethode voor het rekenvaardig schatten. Antwoorden tussen 16^e en 84^e percentiel krijgen twee punten; antwoorden onder het 2^{de} percentiel en boven het 98^e percentiel krijgen geen punten. De maximale score die behaald kan worden is veertig punten. De waarden van de percentielen zijn gebaseerd op de groep zonder tijdslimieten. Om vergelijking mogelijk te maken werden deze waarden eveneens gehanteerd bij de tijdslimietgroep. Antwoorden waarbij de onjuiste eenheid is gehanteerd, zoals ponden bij het domein 'afstand' zijn eveneens als oncorrect beoordeeld. Dit sluit aan bij de beoordeling van Silverberg et al. (2007). Indien door de respondent de meeteenheid niet is aangegeven, heeft de beoordelaar geschat of het antwoord mogelijk in lijn lag met de antwoorden van de andere respondenten. Wanneer dit niet het geval was, is dit antwoord als missing value verwerkt.

Het zou vanzelfsprekend zijn dat wanneer de Biber Cognitive Estimation Test (BCET) gebruikt wordt, dezelfde scoringsmethode wordt gehanteerd. Bullard et al. (2004) beoordeelden de antwoorden binnen het 5^e en 95^e percentiel als correcte respons en gaven deze correcte antwoorden één punt. Het nadeel van deze scoringsmethode is dat er een vrij grote kans bestaat dat respondenten de maximale score van twintig punten behalen. Hierdoor is het lastiger respondenten te vergelijken en verbanden te testen. De scoringsmethode van Axelrod en Millis (1994) maakt het mogelijk respondenten te vergelijken aangezien een driepuntenschaal gehanteerd wordt. De kans op een maximale score is immers met deze scoringsmethode kleiner geworden.

Bij het rekenvaardig schatten is een totale score van tien punten te behalen. Er wordt één punt gegeven wanneer het antwoord binnen de range van het correcte antwoord valt. Na de beoordeling van rekenvaardig schatten en het toepassen van de scoringsmethode bij kwantitatief schatten zijn de analyses uitgevoerd.

§ 4 Resultaten

In deze paragraaf worden de resultaten van de analyses beschreven. Er wordt gestart met de algemene scores op rekenvaardig schatten en kwantitatief schatten. Met behulp van deze scores is het mogelijk

de samenhang tussen rekenvaardig schatten en kwantitatief schatten te berekenen. Daarna wordt de invloed van tijd beschreven. Het derde onderdeel richt zich op de invloed van de persoonlijke karakteristieken. Het resultatengedeelte eindigt met een analyse van de afgebroken testen op basis van persoonlijke karakteristieken.

§ 4.1 Totale score rekenvaardig schatten en kwantitatief schatten

De maximale score die behaald kon worden bij het rekenvaardig schatten was tien punten. De gemiddelde score voor de respondenten met een tijdslimiet was 4,27 punten ($SD = 2,06$). Respondenten zonder een tijdslimiet hadden een hoger gemiddelde van 5,86 punten ($SD = 2,06$).

Bij het kwantitatief schatten was de maximale score 40 punten. Het gemiddelde puntenaantal van respondenten met een tijdslimiet bedroeg 31,80 punten ($SD = 3,92$). De gemiddelde score voor de respondenten zonder de tijdslimiet was 32,85 punten ($SD = 3,49$).

§ 4.2 Aanpak respondenten rekenvaardig en kwantitatief schatten

In de test is gevraagd naar de aanpak van de respondenten om na te gaan in hoeverre respondenten een antwoord hebben geschat of exact hebben uitgerekend. De meeste respondenten hebben alle sommen en vragen geprobeerd te schatten (tabel vier). Uit de frequenties blijkt dat de frequenties van ‘exact berekend’ zich niet laten spiegelen aan de frequenties van ‘geschat’. Wanneer respondenten hebben aangegeven geen enkele som exact te hebben berekend, zouden hetzelfde aantal respondenten aangeven alle sommen te hebben geschat et cetera.

Tabel 4

Frequenties aanpak respondenten

	Rekenvaardig schatten		Kwantitatief schatten		
	Exact berekend	Geschat	Exact berekend	Geschat	
Geen	112	9	Geen	162	5
1-2 keer	58	13	1-4 keer	58	5

3-4 keer	38	29	5-8 keer	7	16
5-6 keer	18	51	9-12 keer	3	16
7-8 keer	6	33	13-16 keer	2	31
Alle sommen	5	102	Alle vragen	5	164

Met behulp van een Pearson's productmomentcorrelatie is berekend in hoeverre de aanpak van respondenten samenhangt met de score op rekenvaardig schatten en kwantitatief schatten. Hieruit blijkt dat er alleen een significante relatie bestaat tussen de mate waarin een som exact is uitgerekend en de score op rekenvaardig schatten ($R = .27, p = .00$). Er is geen significante relatie te leggen tussen de mate waarin het antwoord is geschat en de score op rekenvaardig schatten ($R = -.09, p = .17$). Kwantitatief schatten kent geen samenhang met de mate van exact berekenen ($R = .09, p = .17$) of schatten ($R = -0.2, p = .73$).

§ 4.3 Samenhang rekenvaardig schatten en kwantitatief schatten

Een Pearson's productmomentcorrelatie is berekend om na te gaan of de score op rekenvaardig schatten samenhangt met de score op kwantitatief schatten. Er is gebleken dat er een kleine positieve samenhang bestaat tussen de scores op rekenvaardig schatten en kwantitatief schatten ($R = .204, p < .002$), tweezijdig)

§ 4.4 Invloed van tijdslimiet op rekenvaardig en kwantitatief schatten

De invloed van een tijdslimiet op de score van rekenvaardig schatten en kwantitatief schatten is berekend met behulp van een ongekoppelde t -toets. Uit de Levene's test blijkt dat er sprake is van een gelijke variantie bij zowel rekenvaardig schatten als kwantitatief schatten (rekenvaardig schatten $\rightarrow F = 0,04, p = .84$; kwantitatief schatten $\rightarrow F = 1,54, p = .22$) waardoor de t -toets met aangenomen gelijke variantie (*equal variance assumed*) is uitgevoerd. Hieruit blijkt dat zowel de gemiddelde score op rekenvaardig schatten en de gemiddelde score op kwantitatief schatten tussen beide groepen significant van elkaar verschillen (rekenvaardig schatten $\rightarrow t(235) = -5,97, p < .00$ tweezijdige toetsing; kwantitatief schatten $\rightarrow t(235) = -2,19, p = .03$ tweezijdig). De Cohen's d is uitgerekend om

de effectgrootte te bepalen. Voor rekenvaardig schatten bedraagt deze effectgrootte -0,78 en wordt hiermee als middel groot effect aangeduid (Baarda, De Goede & Van Dijkum, 2007). Kwantitatief schatten heeft een kleine effectgrootte en bedraagt -0,29. De verschillende groepen kunnen voor 13% de verschillen in score bij het rekenvaardig schatten verklaren, 2% bij het kwantitatief schatten.

§ 4.5 Invloed onderwijsniveau rekenvaardig schatten

Om te bepalen of onderwijsniveau samenhangt met de score op het rekenvaardig schatten is de Spearman's rangcorrelatie berekend. Er blijkt een kleine positieve samenhang te bestaan tussen onderwijsniveau en de score bij het rekenvaardig schatten ($r_s = .29$, $p < .00$, tweezijdig).

§ 4.6 Invloed van persoonlijke karakteristieken op rekenvaardig en kwantitatief schatten

Er is nagegaan of de score op rekenvaardig schatten en kwantitatief schatten kan worden voorspeld door geslacht, leeftijd en onderwijsniveau. Een multiple regressie is uitgevoerd met onderwijsniveau als dummy variabele. Field (2005) meent dat wanneer er dummy variabelen aangemaakt worden, de variabelen die het meest vertegenwoordigd wordt door de onderzoeksgroep als vergelijkingsvariabele dient te worden gehanteerd. Het betreft hier met onderwijsniveau niet om een categorisch kenmerk maar om een variabele op ordinaal niveau, waardoor 'wo' als vergelijkingsvariabele is gehanteerd, ook al werd hbo het meest vertegenwoordigd.

Er gelden verschillende voorwaarden waar idealiter aan voldaan moet worden voordat een multiple regressie kan worden uitgevoerd. De variabelen zijn kwantitatief, categorisch of interval van niveau en de voorspellende variabelen hebben enige variantie. De analyse wees uit dat de voorspellende variabelen niet sterk samenhangen. De toleranties op alle variabelen liggen boven .2, waardoor deze voorwaarde is aangenomen. Een andere voorwaarde voor multiple regressie is dat de voorspellende variabelen niet mogen samenhangen met externe variabelen. In deze multiple regressie worden alle variabelen als voorspellende variabelen gebruikt, waardoor niet is na te gaan of deze variabelen mogelijk samenhangen met externe variabelen. Ten slotte is de spreiding van residuen per X-waarde niet gelijk want bij alle onafhankelijke variabelen is er sprake van een ongelijke verdeling

aan respondenten. Aangezien aan de meeste voorwaarden voldaan werd, is er een multiple regressie uitgevoerd.

Uit de multiple regressie (tabel vijf) blijkt dat de onafhankelijke variabelen voor 13% ($p < .00$) de score op het rekenvaardig schatten voorspellen. De variabelen geven geen significante voorspelling op de score van kwantitatief schatten.

Tabel 5

Multiple regressie rekenvaardig schatten en kwantitatief schatten

	Rekenvaardig schatten			Kwantitatief schatten		
	<i>B</i> *	<i>SE B</i>	β	<i>B</i> **	<i>SE B</i>	β
Constant	6,47	0,53		31,83	0,95	
Geslacht	0,33	0,28	,073	-0,27	0,50	-,035
Leeftijd	-0,001	0,01	-,037	0,04	0,02	,126
Dlbo/Mavo	-3,11	0,68	-,355***	-2,19	1,21	-,147
DHavo/Vwo	-1,44	0,48	-,281***	-0,69	0,85	-,080
DMbo	-1,95	0,50	-,359***	-1,35	0,88	-,146
DHbo	-1,04	0,45	-,226***	0,01	0,80	,002

* $R^2 = .13$; $\Delta R^2 = .10$ ($p < .000$).

** $R^2 = .04$; $\Delta R^2 = .01$ ($p = .18$).

*** $p < .05$.

Een MANOVA zou een volgende analyse kunnen betekenen. Bij een MANOVA is het mogelijk meerdere afhankelijke variabelen te betrekken wanneer men een relatie tussen deze variabelen verwacht (Field, 2005), in dit geval rekenvaardig schatten en kwantitatief schatten. Uit de Pearson's productmomentcorrelatie is naar voren gekomen dat er een kleine samenhang is tussen rekenvaardig schatten en kwantitatief schatten. Uit de multiple regressie is echter naar voren gekomen dat de onafhankelijke variabelen zoals geslacht, leeftijd en onderwijsniveau de score op het kwantitatief schatten niet significant kunnen voorspellen. Op basis van deze uitkomst zal een MANOVA geen significante voorspelling kunnen doen. Een MANOVA is uitgevoerd om deze verwachting te testen. Uit de Box's test voor gelijkwaardige covariantie blijkt dat aan de assumptie voor homogeniteit wordt

voldaan (Box's $M = 70,86$, $F(48/1741,56) = 1,04$, $p = ,393$). Indien geslacht, leeftijd en onderwijsniveau als factoren bij de MANOVA worden betrokken, blijken deze persoonlijke karakteristieken geen significante invloed te hebben op rekenvaardig schatten en kwantitatief schatten (Pillai's Trace = .19, $F = 1,24$, $p = .24$). Wanneer de combinatie geslacht – onderwijsniveau of leeftijd – onderwijsniveau als combinaties van factoren worden gezien, blijken deze combinaties wel een significante invloed te hebben (Pillai's Trace (geslacht-onderwijsniveau) = .17, $F = 2,27$, $p = .02$; Pillai's Trace (leeftijd-onderwijsniveau) = .87, $F = 1,45$, $p = .01$). Uit de ANOVA komt naar voren dat onderwijsniveau effect heeft op rekenvaardig schatten ($F(6) = 3,01$, $p < .01$, $MS = 11,58$). Dit sluit aan bij de Spearman's rangcorrelatie. Daarnaast blijkt er een interactie-effect te bestaan tussen geslacht en leeftijd voor rekenvaardig schatten ($F(18) = 1,88$, $p = .03$, $MS = 7,24$). Tenslotte is er een interactie-effect van leeftijd en onderwijsniveau op kwantitatief schatten ($F(51) = 1,60$, $p = .03$, $MS = 17,65$).

§ 4.7 Persoonlijke karakteristieken en afbreken vragenlijst

Aangezien er een groot aantal afgebroken testen zijn, is het interessant om na te gaan of de tijdslimiet een invloed heeft op het afbreken van de test. Er waren 100 respondenten uit de tijdslimietgroep die halverwege de test zijn gestopt. Bij de groep zonder tijdslimieten waren dat er 53 respondenten.

Van de 100 respondenten uit de tijdslimietgroep hebben 14 respondenten de laatste som van rekenvaardig schatten beantwoord. Uit de groep zonder tijdslimiet waren dit vijf respondenten van de 53. Twaalf respondenten uit de tijdslimietgroep hebben het derde gedeelte van de test, kwantitatief schatten, gestart. Bij de groep zonder tijdslimiet heeft één respondent dit gedeelte gestart. Uit deze gegevens lijkt het dat respondenten de test zonder tijdslimiet eerder stoppen, welke significant blijkt te zijn ($Chi^2 = 7,46$; $df = 1$; $p < .006$; $n = 389$). De mate waarin de tijdslimiet invloed heeft op het voltooien van de test is aan te geven met behulp van een Cramér's V, welke bedraagt 0,14 ($p < .006$).

Met het uitrekenen van een Chi-kwadraat is het eveneens mogelijk om verschillen tussen het geslacht en het voltooien van de vragenlijst na te gaan. Hieruit blijkt dat het geslacht geen significant verschil uit maakt bij het voltooien of afbreken van de test ($Chi^2 = 3,01$, $df = 1$; $p = .08$; $n = 385$).

Er is tevens nagegaan of verschillen tussen de respondenten op basis van onderwijsniveau significant zijn. Hiervoor is een Mann-Whitney U-toets uitgevoerd, aangezien het hier gaat om een test

variabele van ordinaal niveau. Uit deze analyse blijkt dat het voltooien van de test niet verschilt tussen de diverse onderwijsniveaus ($U = 15958,50$, $p = .45$, $n = 378$).

Ten slotte is nagegaan met behulp van een t -toets of leeftijd een significant verschil uitmaakt voor het voltooien of afbreken van de test. Uit de Levene's test blijkt dat er geen sprake is van gelijke variantie waardoor de t -toets voor ongelijke variantie is uitgevoerd (*equal variances not assumed*; $F = 7,50$; $p < .01$). De t -toets geeft aan dat leeftijd geen verschil uitmaakt voor het voltooien van de test ($t(274,44) = -.95$, $p = .35$).

§ 5 Conclusie

De centrale vraag bij dit onderzoek is: *hangt het kwantitatief schatten samen met rekenvaardig schatten en in welke mate beïnvloeden de factoren tijdslimiet en persoonlijke karakteristieken de schattingsvaardigheden?* Aan deze onderzoeksvraag zijn enkele deelvragen gekoppeld, welke door de uitgevoerde analyses beantwoord kunnen worden.

In dit onderzoek werd nagegaan of rekenvaardig schatten samenhangt met kwantitatief schatten. Uit de analyse is naar voren gekomen dat er sprake is van een kleine en positieve samenhang. Wanneer iemand een hogere score heeft het rekenvaardig schatten, bestaat er een kleine kans dat er een hoge score voor kwantitatief schatten gerealiseerd wordt. Deze bevinding is een aanvulling op de bevindingen van Siegler en Booth (2004) dat numeriek schatten samenhangt met rekenkundige bewerkingen.

De tweede deelvraag betreft het nagaan van de invloed van een tijdslimiet op rekenvaardig en kwantitatief schatten. De factor tijdslimiet heeft een negatieve invloed op rekenvaardig schatten en kwantitatief schatten. Bij zowel rekenvaardig schatten als kwantitatief schatten, hebben respondenten uit de groep zonder tijdslimiet een significant hogere score. Rekenvaardig schatten kent het regelprincipe en is daardoor niet geschikt voor onbewuste overwegingen (Dijksterhuis & Nordgen, 2006). Deze negatieve invloed van tijd op rekenvaardig schatten is echter niet in lijn met het onderzoek van Rhymer et al. (2002). Daarnaast is nagegaan welke invloed tijd heeft op het maken van kwantitatieve schattingen. Het kan mogelijk niet geschikt zijn om alle aspecten van de schattingssituatie te overwegen. Tijdslimiet heeft in dit onderzoek echter geen significante invloed. De

effectgrootte is het grootst maar gemiddeld bij rekenvaardig schatten (13%) en zeer klein bij het kwantitatief schatten (2%). Of bewuste overwegingen dan wel bewuste beslissingen geschikter zijn voor het maken van schattingen kan door de kleine effectgrootte bij de *t*-toets nog niet worden bepaald. De factor tijdslimiet heeft ten slotte een invloed gehad op het doorlopen dan wel afbreken van de vragenlijst. Respondenten uit de groep zonder tijdslimieten stopten eerder met de vragenlijst dan de respondenten uit de tijdslimietgroep.

Ten slotte is onderzocht of persoonlijke karakteristieken invloeden hebben op het maken van rekenvaardige schattingen en kwantitatieve schattingen. Onderwijsniveau kent een kleine positieve significante relatie met rekenvaardig schatten. Respondenten met een hoger onderwijsniveau maakten het gedeelte van rekenvaardig schatten beter dan de respondenten met een lager onderwijsniveau. Andere karakteristieken zoals geslacht of leeftijd heeft geen invloed op de score van rekenvaardig schatten. De persoonlijke karakteristieken geslacht, leeftijd en onderwijsniveau hebben geen invloed op het maken van kwantitatieve schattingen. Deze bevinden sluiten aan bij Gillepsie et al. (2004) en Spencer en Johnson-Greene (2009) op het gebied van geslacht en leeftijd. Op het gebied van onderwijsniveau spreekt deze bevinding zich tegen met de onderzoeken van Della Sala et al. (2003) en Spencer en Johnson-Greene (2009).

De meeste respondenten hebben geprobeerd vrijwel alle antwoorden te schatten bij zowel rekenvaardig schatten als kwantitatief schatten. Bij rekenvaardig schatten is er echter meer spreiding van de frequenties dan bij kwantitatief schatten. Met behulp van kwantitatief onderzoek is niet na te gaan of rekenkundige bewerkingen een strategie is voor het maken van kwantitatieve schattingen. Hierdoor kan nog niet geconcludeerd worden dat deze strategie mogelijk een aanvulling is op het Metrics and Mappingsmodel van Siegler en Brown (1993).

§ 6 Discussie

Dit onderzoek heeft opgeleverd dat er een relatie is ontdekt tussen rekenvaardig schatten en kwantitatief schatten. Het uitvoeren van rekenkundige berekeningen kan als mogelijke strategie voor kwantitatief schatten nader worden onderzocht. Daarnaast is naar voren gekomen dat een tijdslimiet een negatieve invloed heeft op het genereren van antwoorden op rekenvaardig schatten en mogelijk

ook op kwantitatief schatten. Dat presteren onder tijdsdruk gevolgen kan hebben op de prestaties in het algemeen, wordt met deze bevindingen nogmaals bevestigd. Aangezien het bij dit onderzoek gaat om een pioniersonderzoek, zijn er een aantal beperkingen en aanbevelingen bij dit onderzoek te plaatsen.

Veel respondenten hebben de test afgebroken waardoor een grote bron van informatie ontbreekt. Uit de afgebroken testen blijkt dat het grotendeel van de respondenten gestopt zijn bij het rekenvaardig schatten. De meeste afgebroken testen waren eveneens bij de tijdslimietgroep. Opvallend is dat respondenten uit de groep zonder tijdslimiet de vragenlijst significant eerder hebben beëindigd. Een reden voor het afbreken van de vragenlijst kan zijn dat men geen vertrouwen heeft in zijn of haar capaciteit en dat men schrikt van de sommen. De meeste respondent zijn immers niet gewend om schattend te rekenen, omdat deze strategie in het realistische rekenonderwijs pas sinds de laatste decennia wordt onderwezen. De respondenten die de test geheel doorlopen hebben, beheersen mogelijk over een betere rekenvaardigheid en beschikken over betere strategieën met betrekking tot rekenvaardig schatten. Wanneer respondenten die hebben afgehaakt de vragenlijst wel hadden afgemaakt, zodat deze testen betrokken konden worden bij de analyses, hadden er mogelijk andere of sterkere relaties gelegd kunnen worden. De verschillen in scores tussen de twee onderzoeksgroepen op het gebied van rekenvaardig schatten en kwantitatief schatten waren dan wellicht anders geweest. Dit verschil kan zich resulteren in een ander samenhang tussen rekenvaardig schatten en kwantitatief schatten. Een andere mogelijke reden voor de afgebroken testen heeft te maken met de technische mogelijkheden. De test is op een internetsite gepubliceerd en vraagt een goede internetverbinding om deze test te kunnen doorlopen. Drie respondenten hebben aangegeven dat de internetpagina halverwege de test niet meer werd geladen en als gevolg hiervan werd afgebroken. Deze situatie heeft zich bij andere respondenten mogelijk ook voorgedaan. Het is dan ook een beperking van deze onderzoeksdesign. Onderzoek met behulp van een ander design kan dit mogelijk ondervangen, al kennen andere designs andere nadelen. Het afnemen van de test op papier is bijvoorbeeld minder veilig voor de respondenten en zij zullen minder snel bereid zijn om mee te werken of maken sneller onnodige fouten. Ten slotte vroeg dit design veel van de computervaardigheden van de respondenten. Wanneer iemand niet over de computervaardigheden of typvaardigheden beschikt, is de kans groot om de test te stoppen.

De respondenten die bij dit onderzoek betrokken zijn, worden gekenmerkt door een diversiteit aan leeftijden. Dit is een beperking van het onderzoek omdat leeftijd mogelijk een invloed heeft op het maken van kwantitatieve schattingen. Volgens Della Sala et al. (2004) kunnen mensen naarmate men ouder wordt betere kwantitatieve schattingen maken. Dit zou kunnen betekenen dat wanneer de gemiddelde leeftijd bij onderzoeken hoger ligt, dit zou kunnen leiden tot betere kwantitatieve schattingen. Echter de onderzoeksgroep in dit onderzoek is te klein om dit verschil te onderzoeken. Een vervolgonderzoek zal zich kunnen richten op een specifieke leeftijdscategorie. Interessant is hierbij of tieners onder de zestien jaar inderdaad meer fouten maken en een lagere score realiseren dan (jonge) volwassenen. Uitkomsten uit dergelijk onderzoek kan de bewering van Bacon (zoals geciteerd in Liss et al., 2000) bevestigen of tegenspreken. Bacon beweert dat schattingsvaardigheden zich tijdens de late tienerjaren beginnen te ontwikkelen en dat deze vaardigheden zich door de jaren heen verbeteren.

De andere diversiteit binnen de onderzoeksgroep is op het gebied van onderwijsniveau. Met dit design, het publiceren van de test op een internetsite, is er nauwelijks invloed uit te oefenen op de verspreiding van de test. Doordat respondenten de link naar de internetsite hebben doorgestuurd zijn er veel respondenten bereikt, maar is er een ongelijke verdeling aan leeftijden en onderwijsniveau ontstaan. Uit dit onderzoek blijkt dat onderwijsniveau van invloed is op rekenvaardig schatten en niet op kwantitatief schatten, waarbij deze bevinding in tegenspraak is met eerdere onderzoeken in de wetenschappelijke literatuur. Vervolgonderzoek zou zich kunnen richten op één specifiek onderwijsniveau. Een andere mogelijkheid is het betrekken van meer respondenten per onderwijsniveau om verschillen in antwoorden en strategieën te achterhalen. Wanneer er meer respondenten worden betrokken of minder sprake is van spreiding, heeft onderwijsniveau mogelijk een grotere invloed op het maken van kwantitatieve schattingen.

Onderzoek naar verschillende tijdslimieten is eveneens nodig. Bij dit onderzoek is een tijdslimiet van vijftien seconde gehanteerd, gebaseerd op de pilot. Mogelijk is vijftien seconde om antwoord te geven op de schattingsvragen onrealistisch en onhaalbaar. Men heeft tijd nodig om de vraag te lezen en antwoord te geven. Een te korte tijd heeft invloed op het goed lezen van de vragen of

op de interpretatie van de vragen. Er zijn echter een groot aantal volledige testen en hebben veel respondenten wel antwoord kunnen geven binnen deze tijdslimiet.

In het meetinstrument zijn enkele schattend rekensommen opgenomen. Deze sommen komen uit lesmethoden voor leerlingen uit groep acht. De range van een goed antwoord is gebaseerd op de afrondstrategie op beide getallen. Mogelijk zijn de gehanteerde ranges te nauw, zodat minder respondenten een punt kregen toegekend. Het andere uiterste is ook mogelijk, dat de range te breed was en hierdoor te veel respondenten een punt hebben ontvangen. De vraag bij het rekenvaardig schatten is: wat is een goed antwoord? Opmerkelijk is dat de lesmethoden deze goede antwoorden niet geven, aangezien de leerlingen het goede antwoord alsnog exact moeten uitrekenen. Onderzoek het goede antwoord bij rekenvaardig schatten, is nodig.

Voor het gedeelte van kwantitatief schatten is gebruik gemaakt van de Biber Cognitive Estimation Test [BCET] van Bullard et al. (2004). Naar aanleiding van dit onderzoek kan men vraagtekens zetten bij de betrouwbaarheid en validiteit van deze test naar kwantitatief schatten. Het accepteren van items in de definitieve versie van de BCET en de scoringsmethode zijn beide dubieus te noemen. De validiteit is bijvoorbeeld niet gemeten met behulp van een factoranalyse en de scoringsmethode van drie standaarddeviaties maakt het vergelijken tussen respondenten haast onmogelijk. Voor dit onderzoek was het niet de doelstelling om de items van de BCET en de scoringsmethode op betrouwbaarheid en validiteit te testen. Dit soort onderzoeken vragen immers om een pre- en posttest totdat het gewenste niveau van betrouwbaarheid is behaald. Wel zijn enkele items in dit onderzoek gedeeltelijk aangepast om deze items te verduidelijken. Onspecifieke en onduidelijke items kunnen bij respondenten verwarring brengen. Het is de vraag of deze aanpassingen een positieve bijdrage hebben geleverd aan de duidelijkheid van de items. Het hanteren van de scoringsmethode van Axelrod en Millis (1994, zoals geciteerd in Silverberg et al., 2007) realiseert een grotere variatie tussen respondenten, waardoor het mogelijk tot betere vergelijkingen en uitspraken te komen.

Het tegenstrijdige met kwantitatieve schattingstesten is de afwezigheid van het goede antwoord. Items in dergelijke testen mogen immers geen vragen bevatten waarvan “het exacte antwoord niet direct beschikbaar is” (Brand et al., 2003, p. 575). De discussie welke door onderzoekers binnen dit domein gevoerd dient te worden, richt zich op de vraag: hoe kunnen

respondenten worden beoordeeld op hun schattingsvaardigheden wanneer niet bekend is wat goed is? Indien schattingsvaardigheden worden beoordeeld door middel van een vergelijking met een controlegroep, is er een grote kans om beoordelingsfouten te maken. Bij een slecht presterende controlegroep worden respondenten uit de vergelijkingsgroep ten onrechte als zwakke schatters bestempeld. Della Sala et al. (2003) hebben geprobeerd dit aspect te ondervangen door een correcte respons te hanteren en vanuit deze correcte respons een range toe te passen, waarbinnen het antwoord als goed wordt beoordeeld. Het is echter onduidelijk hoe Della Sala et al. (2003) aan deze correcte antwoorden zijn gekomen. Forrester en Latham (1990) geven aan dat het goede antwoord een *'best-fit'* antwoord is waarbij rekening wordt gehouden met belangrijke achtergrondfactoren. Het is echter de vraag onder welke omstandigheden een antwoord *'best-fit'* is en of deze achtergrondfactoren voor de schatter ook bekend en relevant zijn.

Siegler en Brown (1993) hebben met het Metrics and Mappings model een begin gemaakt om strategieën die men in kan zetten bij het kwantitatief schatten te ordenen. Het laatste discussiepunt is dan ook een aanbeveling gericht op onderzoek naar de strategieën die men gebruikt om op een schattingsvraag antwoord te geven. In dit onderzoek hebben respondenten aangegeven het antwoord niet exact uit te rekenen. Maar zijn de respondenten zich er daadwerkelijk van bewust dat zij het antwoord exact proberen uit te rekenen? Schatten is immers zo sterk verweven in het dagelijks leven van mensen, dat dit een onbewuste strategie geworden kan zijn. Daarnaast bleek uit de aanpak van de respondenten dat het verschil waarin een som of vraag exact werd uitgerekend of werd geschat, niet geheel overeenkwam. Men kan zich afvragen of de respondenten op het gebied van schatten, sociaal wenselijk heeft geantwoord. Uit kwalitatief onderzoek kunnen andere strategieën naar voren komen welke in het model van Siegler en Brown nog niet zijn opgenomen. Uit dit onderzoek is gebleken dat rekenvaardig schatten niet alleen samenhangt met kwantitatief schatten, maar mogelijk een strategie kan zijn om te komen tot kwantitatieve schattingen.

Referenties

- Baarda, D.B., De Goede, M.P.M., & Van Dijkum, C.J. (2007). *Basisboek Statistiek met SPSS. Handleiding voor het verwerken en analyseren van en rapporteren over (onderzoeks)gegevens*. Houten: Wolters-Noordhoff Groningen.
- Barabassy, A., Beinhoff, U., & Riepe, W. (2010). Cognitive estimation in aged patients with major depressive disorder. *Psychiatry Research*, *176*(1), 26-29.
- Beishuizen, M., Van Putten, C.M., & Van Mulken, F. (1997). Mental arithmetic and strategy use with indirect number problems up to one hundred. *Learning and Instruction*, *7*(1), 87-106.
- Brand, M., Kalbe, E., Fuijware, E., Huber, M., & Markowitsch, H. (2003). Cognitive estimation in patients with probable Alzheimer's disease and alcoholic Korsakoff patients. *Neuropsychologia*, *41*(5), 575-584.
- 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.
- 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.
- Della Sala, S., MacPherson, S.E., Philips, L.H., Sacco, L., & Spinnler, H. (2004). The role of semantic knowledge on the cognitive estimation task. Evidence from Alzheimer's disease and healthy adult aging. *Journal of Neurology*, *251*(2), 156-164.
- 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*(17), 1005-1007.
- Dijksterhuis, A., & Nordgren, L.F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science*, *2*(1), 95-109.
- Field, A. (2005). *Discovering Statistics Using SPSS*. Londen: SAGE Publications.
- Forrester, M.A., & Latham, J. (1990). Exploring estimation in young primary school children. *Educational Psychology*, *10*(4), 283-301.

- Gillespie, D.C., Evans, R.I., Gardener, E.A., & Bowen, A. (2002). Performance of older adults on test of cognitive estimation. *Journal of Clinical and Experimental Neuropsychology*, *24*(3), 286-293.
- Huitema, S., Van de Klis, A., Timmermans, M., & Erich, L. (2001). *Wereld in getallen. Groep 8;boek B*. 's Hertogenbosch: Malmberg.
- Klunder, R. (2005). *Schattend rekenen in het primair onderwijs*. Ongepubliceerde masterthesis, Universiteit Twente, Enschede, Nederland.
- 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.
- Limesurvey (2010). *Limesurvey, de open source survey application*. Gevonden op 19 april 2010, op: <http://www.limesurvey.org>
- Liss, M., Fein, F., Bullard, S., & Robins, D. (2000). Brief report: Cognitive estimation in individuals with pervasive developmental disorders. *Journal of Autism and Development*, *30*(6), 613-618.
- Murray, K.B., & Brown, N.R. (2009). A feature-based inference model of numerical estimation: The split-seed effect. *Acta Psychologica*, *131*(3), 221-234.
- 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.
- Rubenstein, R.N. (1986). Varieties of estimation. In Schoen, H.L., & Zweng, M.J. (Ed.), *Estimation and Mental Computation*. (pp. 157-162). Virginia: The National Council of Teachers of Mathematics, INC.
- Shallice, T., & Evans, M.E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex*, *14*, 294-303.
- Silverberg, N.D., Hanks, R.A., & McKay, C. (2007). Cognitive estimation in traumatic brain injury. *Journal of International Neuropsychological Society*, *13*, 898-902.
- Siegler, R.S., & Opfer, J.E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*(3), 237-243.

Smith, M.L., & Milner, B. (1984). Differential effect of frontal-lobe lesions on cognitive estimation and spatial memory. *Neuropsychologica*, 22(6), 697-705.

Spencer, R.J., & Johnson-Greene, D. (2009). The cognitive estimation test (CET): Psychometric limitations in neurorehabilitation populations. *Journal of Clinical and Experimental Neuropsychology*, 31(3), 373-377.

Star, J.R., & Rittle-Johnson, B. (2009). It pays to compare: An experimental study on computational estimation. *Journal of Experimental Child Psychology*, 102(4), 408–426.

Vijgen, N., & Van Broekhoven, H. (2000). *Pluspunt Plusboek*. 's Hertogenbosch: Malmberg.

Vijgen, N., & Van Broekhoven, H. (2000). *Pluspunt Opdrachtenboek*. 's Hertogenbosch: Malmberg.

Von Helversen, B., & Rieskamp, J. (2008). The mapping model: A cognitive theory of quantitative estimation. *Journal of Experimental Psychology*, 137(1), 73-96.

Appendix A

Meetinstrument

Welkom

Welkom bij de test voor schattingsvaardigheden, de test die ik voor mijn onderzoek gebruik. Weet u hoeveel lucifers er in een doosje zitten? Of hoe hoog een stoplicht hangt? Waarschijnlijk weet u het antwoord niet, maar zult u het antwoord schatten. Ik voer een onderzoek uit naar de factoren die het maken van schattingen beïnvloeden zoals tijd of rekenen. Ik hoop dat u mij daarbij wilt helpen. Deze test bestaat uit drie delen. Bij het eerste gedeelte worden er enkele persoonlijke gegevens van u gevraagd. Het tweede gedeelte bestaat uit een aantal rekensommen waarbij u antwoord mag geven door te schatten. De test wordt afgesloten met 20 schattingsvragen. Het invullen van de test zal maximaal 10 minuten duren. Ik wens u veel plezier en succes! Ik wil u alvast bedanken voor het invullen van deze test!

Persoonlijke gegevens

Voordat de test begint, worden er enkele persoonlijke gegevens verzameld.

Geslacht:

- man
- vrouw

Leeftijd:

Hoogst genoten opleiding:

- lbo
- mavo
- havo
- vwo
- mbo
- hbo
- wo

Schattend rekenen

Bij dit gedeelte van de test krijgt u 10 sommen om te beantwoorden. Het exacte antwoord hoeft niet gegeven te worden, maar probeer het zo goed mogelijk te schatten. Typ het antwoord in het balkje

onder de som. U mag tijdens het beantwoorden van de som geen gebruik maken van een rekenmachine of kladpapier. (Let op: u heeft voor elke som 15 seconde de tijd. Wanneer de tijd voorbij is, kunt u geen antwoord meer geven). U krijgt eerst een voorbeeld som

- $797 - 249$
- $5,15 \times 12,93$
- $418 : 8$
- $3487 + 2829$
- $87,15 - 19,4 - 0,88$
- $51 \times 12,992$
- $4835 + 975 + 1214$
- $6,18 + 12,57 + 19,38$
- 48×72
- $8717 : 23$

Aanpak schattend rekenen

Hoe vaak heeft u bij dit gedeelte een som exact uitgerekend?

- Geen
- 1-2 keer
- 3-4 keer
- 5-6 keer
- 7-8 keer
- alle sommen

Hoe vaak heeft u bij dit gedeelte het antwoord geschat?

- Geen
- 1-2 keer
- 3-4 keer
- 5-6 keer
- 7-8 keer
- alle sommen

Schatten

Dit is het laatste gedeelte van de test. In dit gedeelte geeft u antwoord op 20 schattingsvragen. Er is een grote kans dat u het exacte antwoord niet weet, dus probeer een goede schatting te geven. U kunt het antwoord typen in het balkje onder de vraag. Het antwoord mag uit één schatting bestaan en mag geen range bevatten (bijvoorbeeld tussen 125-150 meter). Geef hierbij ook aan om welke meeteenheid het gaat, bijvoorbeeld gram, meters of seconde. Dit mag u afkorten, bijvoorbeeld 'kg', 'sec' of 'm'. (Let op: u heeft per vraag 15 seconden de tijd om antwoord te geven. Wanneer de tijd voorbij is, kunt u geen antwoord meer geven.).De test zal beginnen met een voorbeeldvraag.

Voorbeeld vraag:

Hoe lang is een wagon van een trein?

Het antwoord is 25 meter.

Nu begint de schattingstest. Succes!

- Hoeveel pitten zitten er in een watermeloen?
- Hoeveel weegt een mobiele telefoon?
- Hoeveel stokjes spaghetti zitten er in een pak van één pond?
- Wat is de afstand dat een volwassene in een avond kan lopen?
- Hoe hoog kan iemand op een trampoline springen?
- Hoe lang duurt het bouwen van een woonhuis?
- Hoeveel wegen 12 normale appels bij elkaar?
- Wat is de afstand dat een paard en wagen in één uur kan afleggen?
- Hoeveel keren kan je je tanden poetsen van een tube tandpasta?
- Hoeveel chips zitten er in een gewone zak van 200 gram?
- Hoelang duurt het voor een volwassene om één kant van een A4 vol te schrijven?
- Wat is de leeftijd van de oudste, levende persoon in Nederland?
- Hoe lang is een eetlepel?
- Hoeveel weegt een inklapbare tuinstoel?
- Hoeveel tijd kost het om een T-shirt te strijken?
- Hoe lang is de nek van een giraffe?
- Hoeveel plakken brood zitten er in een brood van 500 gram?
- Hoeveel weegt een paar mannenschoenen?
- Hoe zwaar is een paard?
- Hoe lang duurt het voordat verse melk na de houdbaarheidsdatum in de koelkast zuur wordt?

Aanpak schatten

In hoeverre heeft u bij dit gedeelte het antwoord geprobeerd uit te rekenen?

- Geen
- 1-4 keer
- 5-8 keer
- 9-12 keer
- 13-17 keer
- alle vragen

In hoeverre heeft u de antwoorden van dit gedeelte geschat?

- Geen
- 1-4 keer
- 5-8 keer
- 9-12 keer
- 13-17 keer
- alle vragen