DISSECTIE EN OBSERVATIE

VISUEEL TOETSEN VAN VISUEEL LEREN

Kitty van Gendt

PROMOTIE COMMISSIE

<i>Voorzitter</i>	Prof. dr. ir. A. J. Mouthaan ■ Universiteit Twente
<i>Promotoren</i>	Prof. dr. ir. P. W. Verhagen ■ Universiteit Twente Prof. dr. C. A. W. Glas ■ Universiteit Twente
<i>Leden</i>	Prof. dr. J. J. H. van den Akker ■ SLO Prof. dr. ir. T. J. H. M. Eggen ■ Universiteit Twente Prof. dr. W. A. J. M. Kuiper ■ Universiteit Utrecht Prof. dr. J. M. Pieters ■ Universiteit Twente Prof. dr. A. J. W. Scheurink ■ Rijksuniversiteit Groningen

Gendt, K. van
Dissectie en observatie. Visueel toetsen van visueel leren

Proefschrift Universiteit Twente, Enschede.
ISBN 978-90-365-3745-2
DOI 10.3990/1.9789036537452

Printer: Ipskamp Drukkers B.V. Enschede
Omslag: Kitty van Gendt
© Copyright, 2014, K. van Gendt

DISSECTIE EN OBSERVATIE
VISUEEL TOETSEN VAN VISUEEL LEREN

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. H. Brinksma,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op donderdag 27 November 2014 om 16:45 uur

door

Kitty Angelique Faridah van Gendt

Geboren op 2 januari, 1971

te Paramaribo, Suriname

Promotoren

Prof. dr. ir. P. W. Verhagen

Prof. dr. C. A. W. Glas

Dit proefschrift is goedgekeurd door de promotoren.

INHOUDSOPGAVE

LIJST VAN FIGUREN EN TABELLEN	iii
VOORWOORD	vii
1. INTRODUCTIE ONDERZOEK MET ONDERZOEKSVRAGEN VANUIT EEN ONTWERPGERICHTE ONDERZOEKSAANPAK	1
1.1 De aanleiding voor het onderzoek	1
1.2 Het onderzoeksgebied	2
2. VISUEEL LEREN EN TOETSEN	19
2.1 Toelichting op de vraagstelling en de aanpak van beschikbare literatuur	19
2.2 Cognitieve processen en de rol van betrokken representaties	23
2.3 Instructies met representaties (visueel leren)	40
2.4 Visuele representaties in relatie tot visueel toetsen	50
2.5 Te onderzoeken vraagstellingen naar aanleiding van de literatuur	59
3. EVALUATIE EN BIJSTELLING VAN DE INSTRUCTIE	63
3.1 Toelichting op evaluatie bestaande instructie	63
3.2 1ste analyse Anatomie rat	64
3.3 De toets	74
3.4 De instructie	77
3.5 Resultaten	91
3.6 Discussie en conclusie	96

4. CONSTRUCTIE VAN DE TOETS	99
4.1 Analyse van de 1 ^{ste} toets: papierenversie	99
4.2 (Her)definiëring van het construct voor dit onderzoek	104
4.3 Vervolg onderzoek	119
5. VISUELE CBT-TOETS – 1E EXPERIMENT	121
5.1 Introductie	121
5.2 Psychometrische werkwijze	122
5.3 Methoden	126
5.4 Resultaten	128
5.5 Discussie en conclusies	139
5.6 Bijstelling van de toets	140
6. VISUELE CBT-TOETS – 2E EXPERIMENT	149
6.1 Introductie	149
6.2 Het Experiment	149
6.3 Resultaten	154
6.4 Discussie en conclusie	183
7. VISUEEL TOETSEN VAN VISUEEL LEREN	191
7.1 Onderzoeksproblematiek en onderzoeksaanpak	191
7.2 Beantwoording van de onderzoeksvragen	193
7.3 Eindresultaat voor de faculteit wiskunde en natuurwetenschappen	198
7.4 Slotopmerking	199
REFERENTIES	201
NEDERLANDSE SAMENVATTING	209
ENGLISH SUMMARY	213
BIJLAGEN	217
Bijlagen horende bij hoofdstuk 3	217
Bijlagen horende bij hoofdstuk 4	233
Bijlagen horende bij hoofdstuk 5	242
Bijlagen horende bij hoofdstuk 6	248

LIJST VAN FIGUREN EN TABELLEN

FIGUREN

1.1	Overzicht onderwijssituatie	2
1.2	Overzicht onderzoekstraject	17
2.1	Overzicht weergaven rattenmaag	28
2.2	Analoge illustratie van een maag	29
2.3	Logische of arbitraire illustratie	30
2.4	Diagram band en treksysteem	33
3.1	Overzicht bestaande en aangepaste 'vernieuwde' instructie	84
3.2	Overzicht experimentele en controle condities	85
3.3	Onderzoeksdesign 1	88
3.4	Onderzoeksdesign 2	88
3.5	Gemiddelde testcores op natoets en retentietoetsen	93
3.6	Gemiddelde testcores op natoets en retentietoetsen	95
4.1	Definitie van het te toetsen construct	106
4.2	Voorbeeld itemformat: Herkennen	109
4.3	Voorbeeld itemformat: Benoemen	110
4.4	Voorbeeld itemformat: Relaties	112
4.5	Visualisatievormen	118
5.1	2*2 Factorieel ontwerp realisme en context	126
5.2	Onderzoeksdesign	126
5.3	Visualisatie conditie 'met context' oorspronkelijk en aangepast	142
5.4	Overzicht representatievormen van te toetsen constructen	143
5.5	Overzicht item in relatie tot dekking construct	146
6.1	Overzicht variabelen realisme en context	150
6.2	Onderzoeksdesign	150
6.3	Voorbeeld van vaardigheid op twee items	167
6.4	Voorbeelden van geclassificeerde reponsen	180

TABELLEN

2.1	Rangschikking representaties	39
2.2	Taxonomie CFR-items	58
2.3	Relatie leerdoeltype en toetstaken	58
3.1	Analyse dissectiepracticum n.a.v. Stakes Countenance	66
3.2	Gemiddelde testcores en standaarddeviaties practicum deelname	91
3.3	Anova-analyse experimentele condities	92
3.4	Gemiddelde testcores en standaarddeviaties op alle toetsen	92
3.5	Repeated measures analyse op drie tijdstippen	93
3.6	Gemiddelde testcores en standaarddeviaties practicum deelname	94
3.7	T-toets met 2 onafhankelijke steekproefen	94
3.8	Gemiddelde testcores en standaarddeviaties op alle toetsen	94
3.9	Repeated measures analyse op drie tijdstippen	95
3.10	Betrouwbaarheidsmetingen	96
4.1	Itemanalyse papierentoets	100
4.2	Overzicht van de resultaten van de interviews	105
4.3	Overzicht van de vragen van de toets	113
4.4	Overzicht itemsformats en toetsvormen	117
5.1a	Overzicht analyse gehele test	129
5.1b	Overzicht analyse toetstaak Herkennen	130
5.1c	Overzicht analyse toetstaak Benoemen	131
5.1d	Overzicht analyse toetstaak Relaties	132
5.2a	Resultaten Mokkenanalyse op toets en deoltoetsen	135
5.2b	Overeenkomsten en verschillen geschaalde items	135
5.3	Correlatiematrices van variabelen herkennen, benoemen en relaties	137
5.4	Onderzoek gelijkheid correlatiematrices	138
6.1a	Overzicht resultaten testanalyse	154
6.1b	Vergelijking toetstaken per toetsversie	155
6.1c	Resultaten itemanalyse toetstaak Herkennen	156
6.1d	Resultaten itemanalyse toetstaak Benoemen	157
6.1e	Resultaten itemanalyse toetstaak Relaties	158
6.2	Vergelijking 1ste CBT toetsversie met de 2de CBT toetsversie	163
6.3	ANOVA: overzicht effecten variabelen op test	165
6.4	MANOVA: overzicht effecten op de toetstaken	166
6.5	IRT analyses	169
6.6	Correlatiematrices variabelen Herkennen, Benoemen en Relaties	171

6.7a	Schatting IRT itemparameters toetstaak Herkennen	172
6.7b	Schatting IRT itemparameters toetstaak Beboemen	173
6.7c	Schatting IRT itemparameters toetstaak Relaties	174
6.8	Overzicht afwijkende en zwakke items	175
6.9	Visualisatievormen van de items met afwijkend gedrag	177
6.10	Totalen voor verwijzingen naar instructievormen	181
6.11	Instructie versus inhoud	182
6.12	Invloed representatievormen	183
6.13a	Herkennen, bruikbare itemvarianten	186
6.13b	Benoemen, bruikbare itemvarianten	187
6.13c	Relaties, bruikbare itemvarianten	188

VOORWOORD

'Its good to have an end to journey toward but it is the journey that matters, in the end.'

Ursula Leguin

Toen dit traject startte was helder welk doel moest worden bereikt evenals de wijze waarop, het bijbehorende tijdpad en wie mij hierin zou vergezellen. Zo logisch maar in tijd zo anders uitgepakt. Het onderzoek reikte verder dan het ontwerpen van een toets en vereiste dat instructie en toets als één geheel werden onderzocht. Velen hebben mij op dit pad vergezeld waarvoor mijn dank aan iedereen en enkelen in het bijzonder.

Het onderzoek vond plaats in Groningen met wetenschappelijke sturing vanuit Twente. Er ging een wereld voor mij open aan de Faculteit Dierfysiologie met indrukwekkende dissectiepractica, zoals het kalfshart en natuurlijk de anatomie van de rat, die blijven verwonderen over de simpelijkende complexiteit van een zoogdier. Met de BIO-cooters -Hens, Solita, Nicole en Eduard- onder het toeziend oog van Niek Meijer en later Anton Scheurink- is veel werk verzet om het onderwijs te verbeteren met gebruikmaking van de nieuwste technologische mogelijkheden. Veel beeldmateriaal is gedigitaliseerd en bewerkt ter minimalisering van proefdiergebruik en ter toetsing van studenten over correcte en niet-correcte anatomische structuren. Vanuit kunstacademie Minerva zijn prachtige anatomische tekeningen geproduceerd door Jeroen. Het was pionieren! Ondertussen heeft de technologie zich voort ontwikkeld met nog meer mogelijkheden die kunnen bijdragen aan een doorontwikkeling van het toetsprogramma. In Twente vond mijn ontwikkeling op wetenschappelijk gebied plaats bij Toegepaste Onderwijskunde met sturing door Pløn Verhagen en later ook Cees Glas waarbij ik met collega's zoals Bert discussieerde over diverse onderzoeken. In de afrondingsfase ondersteunde Sandra Schele in de opmaak van het manuscript waarvoor mijn dank.

Pløn, Niek en Cees, ik ben jullie zeer erkentelijk voor jullie inzet en motivatie om mij niet alleen de kans te geven dit onderzoek uit te voeren maar deze vooral ook af te ronden. Ik vond het enorm fijn met jullie te werken.

Dit alles had ik niet kunnen bereiken zonder de steun van mijn naasten, (schoon)familie en vrienden, die altijd interesse toonden gedurende dit traject. Dank daarvoor en ook voor de bemoedigende woorden in de periode dat het minder ging.

Lieve pap en mam, dank voor al jullie liefde, steun, inzet maar vooral vertrouwen in mij.

Tot slot aan de persoon die mij al enige tijd op elk pad in mijn leven vergezelt en ondersteunt. Romesh, dank voor jouw liefde, de ruimte, de steun en bovenal het vertrouwen. Met jou als team is alles binnen bereik. Lieve Timo en Indi, wat ontzettend fijn dat jullie hierbij zijn. Ons gezamenlijk pad is tot nu en in de toekomst een belevenis!

HOOFDSTUK 1

Introductie onderzoek met onderzoeksvragen vanuit een ontwerpgerichte onderzoeksaanpak

Het hier gepresenteerde onderzoek behelst het ontwerpen van een interventie voor een toetsprobleem. In dit hoofdstuk wordt de toetsproblematiek beschreven, de onderzoeksvraag geformuleerd en de keuze voor ontwerpgericht onderzoek beargumenteerd.

Verder biedt dit hoofdstuk een overzicht van de onderzoeksaanpak ter beantwoording van de onderzoeksvraag.

1.1 DE AANLEIDING VOOR HET ONDERZOEK

De directe aanleiding voor het onderzoek is een toetsprobleem dat zich voordoet in het biologieonderwijs aan de Rijksuniversiteit Groningen. Het betreft practica waarin de dissectie van een zoogdier het studieobject is. In dissectiepractica die worden gegeven aan de faculteit Biologie, is het een vereiste dat studenten de anatomie van het zoogdier leren kennen door het verrichten van de dissectie van een rat. De studenten voeren deze dissectietaak uit gedurende een tweedaags practicum waarin een dode rat laag voor laag wordt ontleed. Bij elke laag is het van belang dat verschillende anatomische structuren worden geobserveerd en geïdentificeerd zoals de verschillende spieren en organen. Het voorzichtig ontleden van een rat biedt de studenten de mogelijkheid om elk onderdeel van de rat nauwkeurig te observeren en op deze wijze kennis op te doen van de anatomische structuren van een zoogdier. Tijdens het practicum tekenen de studenten, als instructiemethode, elke dissectielaag volgens aanwijzingen zoals beschreven in een practicumhandleiding. Deze tekeningen dienen later ook als basis voor het toetsproces. Daarbij gaan de docenten uit van de aanname dat deze tekeningen weergeven in hoeverre een student inzicht heeft in anatomische structuren.

In figuur 1.1 wordt de onderwijssituatie schematisch weergegeven.

Figuur 1.1 Overzicht onderwijssituatie

In deze onderwijssituatie doen zich twee problemen voor. Het eerste probleem is dat de toetsmethode door een lange correctieperiode tijdrovend is. Elke student levert een set van tien tekeningen in. In het geval van 150 studenten worden 1500 tekeningen ingeleverd die door een vijftal docenten en studentassistenten wordt gecorrigeerd en beoordeeld. Dit proces is erg arbeidsintensief en duurt in de praktijk langer dan gewenst want studenten krijgen pas in een veel later stadium van het cursusblok feedback op hun prestaties tijdens de instructie. Het tweede probleem is dat de objectiviteit van de toetsmethode niet is gegarandeerd omdat meerdere personen deelnemen aan het beoordelingsproces, er onzekerheid is over de beoordelingscriteria en de beoordelaarsamenstelling verschilt per practicum met zowel ervaren als onervaren beoordelaars.

Dit onderzoek is gestart om deze problemen het hoofd te bieden met als opdracht een toets te ontwikkelen die (a) aansluit op het leerproces, (b) objectief is en (c) studenten in een relatief kort tijdsbestek van feedback voorziet.

Bij het verstrekken van de opdracht was hierbij meteen duidelijk dat voor het tijdrovend met de hand nakijken van tekeningen een alternatief moet worden gezocht.

1.2 HET ONDERZOEKSGBIED

De context van het onderzoek

In de biologie vormt het bestuderen van objecten uit de natuur een essentieel onderdeel van het vakgebied. De bijbehorende instructie wordt veelal zo ingericht

dat het concrete biologische materiaal het studieobject vormt dat van nabij moet worden geobserveerd. Binnen de biologie bestaat, vanuit onderwijstraditie, de overtuiging dat in dit type onderwijssituaties de instructie het effectiefst is als de studenten de waargenomen visuele kenmerken vast leggen in een tekening (Jones, Reed & Weyers, 1994, 1998; Gaupp, 1904). De toetsproblematiek, die in het hier beschreven onderzoek wordt bestudeerd en waar een interventie voor wordt ontworpen, doet zich voor in die context. Het uitgangspunt dat studenten hun kennis opdoen door het observeren van een realistisch object, concreet biologisch materiaal, blijft daarbij centraal. Dit uitgangspunt zit geworteld in de traditie en onderwijsfilosofie van de biologie mede door de affectieve en haptische doelen die worden nagestreefd in de omgang met dit materiaal. In concrete onderwijssituaties zoals die van het dissectiepracticum voor het bestuderen van de anatomie van de rat geeft het kunnen voelen en ruiken van dood biologisch materiaal een extra dimensie aan de ervaring van het onderwijs en als zodanig ook het leerproces. Dit aspect is voor het hier beschreven onderzoek een gegeven. Het onderzoek richt zich er niet op maar er volgen wel twee randvoorwaarden uit. Hoewel het bij het verwerven van kennis over de anatomie van een zoogdier gaat om cognitieve doelen die langs verschillende wegen zouden kunnen worden bereikt, is de eerste randvoorwaarde dat het dissectiepracticum onderdeel moet blijven van de onderwijssituatie. De tweede randvoorwaarde volgt uit het gelimiteerd gebruik van proefdieren. Voor het oplossen van het toetsprobleem kan een toetsingsituatie waarin het realistische object een rol speelt, vergelijkbaar met de instructiesituatie, niet worden overwogen.

Leren en toetsen van visuele kenmerken

In de meeste onderwijssituaties waarin kennis over de visuele kenmerken van realistische objecten tot de leerstof behoort, maken die objecten niet zelf deel uit van de onderwijssituatie maar worden gerepresenteerd door afbeeldingen. In veel gevallen gaat het om instructiemateriaal in de vorm van tekst en plaatjes. Steeds vaker wordt ook gebruik gemaakt van multimediatechnieken waarbij tekstuele elementen en statische of dynamische visuele representaties worden geïntegreerd ter ondersteuning van het leerproces of het toetsen daarvan (Parshall, Davey, & Pashley, 2000, Mayer, 2005). Er is hierdoor een veelzijdig palet aan visuele representatievormen ontstaan dat varieert van eenvoudige schematische lijntekeningen tot dynamische visualisaties die het realistische object nagenoeg geheel benaderen. Al deze representatievormen zijn statische of dynamische *externe* representaties van het realistische object (Schnotz, 2002; 2005, Hagerty, 2005,

Ainsworth, 2006). De term *extern* geeft weer dat het gaat om de concrete (papieren of multimediale) representatie van een realistisch object ter onderscheiding van de interne representatie: het door de persoon geconstrueerde mentale model van dat object. In hoofdstuk 2 wordt voor het hier gepresenteerde onderzoek een rangschikking van de representatiesoorten gepresenteerd (H2, tabel 2.1).

In de onderwijssituatie die het vertrekpunt vormt voor het onderzoek bestaat het instructiemateriaal uit een realistisch object (de rat) waarvan de (deel-)objectkenmerken (kenmerken van de rat en diens anatomische onderdelen) een integraal deel van de lesstof zijn. Het instructiemateriaal bevat zelf dus geen externe representaties. De als resultaat van de instructiemethode door de studenten geproduceerde tekeningen zijn dat wel.

Onderzoek naar leren en toetsen in dit type onderwijssituaties blijkt schaars in het bijzonder als tekeningen de basis vormen voor de beoordeling van verworven kennis. De Graaff en Verhagen (1997) hebben getracht een toetsmethode te ontwikkelen door studenten zelf hun tekeningen te laten nakijken met een computerprogramma waarbij ze hun tekeningen op een computerscherm moesten reconstrueren. Dit onderzoek wordt in hoofdstuk 2 nader besproken samen met ander onderzoek naar alternatieve toetsvormen met visuele representaties (Dwyer, 1978; Martinez, 1998).

Nadere verdieping in de materie heeft geleid tot de algemene conclusie dat het gebruiken van tekeningen als toetsmiddel, in het hier beschreven onderzoek, ondergeschikt is aan de vraag of de toetsprestatie daadwerkelijk uitdrukt of de persoon de veronderstelde kennis over een bepaald kennisdomein heeft opgedaan. Cronbach en Meehl (1955) omschrijven deze kennis over een bepaald domein als een veronderstelde eigenschap die de te toetsen persoon al dan niet beheerst en gebruikt daarvoor de term “construct”. Een goede constructdefinitie waarin deze veronderstelde eigenschap wordt beschreven is een belangrijk uitgangspunt voor de constructie van een kwalitatief goede toets. Bij het toetsen van kennis over visuele kenmerken behelst de constructdefinitie in dit onderzoek vooral de vraag welke visuele kenmerken van belang zijn om de toets met visualisaties een goede afspiegeling te laten zijn van wat geleerd is van het realistische object. Op dit punt is weinig onderzoek voorhanden dat van belang is voor het leggen van relaties tussen de visuele kenmerken van het reële object die in de onderwijssetting worden geleerd en de vorm van representatieve visualisaties van datzelfde object

die als basis kunnen dienen voor het toetsmateriaal. Het Program of Systematic Evaluation (PSE-methode) van Dwyer (1976 – 2007) vormt een uitzondering. Het PSE-programma is een uitgebreid onderzoeksprogramma dat als algemeen doel had te identificeren welke typen visuele materialen het effectiefst de student ondersteunen in het leerproces gebruikmakend van verschillende ondersteunende instructievormen. Dat dit onderzoeksprogramma is uitgevoerd met meer dan 48000 studenten vanaf de jaren zeventig geeft de uitgebreidheid ervan enigszins weer. Deelnemers aan het onderzoeksprogramma kregen steeds vergelijkbare varianten op dezelfde onderwijsmethode aangeboden bestaande uit een tekst van ca. 2000 woorden over het onderwerp kalfshart. Alle deelnemers kregen een tekst waarin de diverse onderdelen, locaties en functies van het hart werd beschreven waarbij de experimentele groepen werden ondersteund met verschillende typen visuele representaties zoals lijntekeningen, al of niet gedetailleerd, en foto's in zwart-wit zowel als in kleur. Het materiaal werd eveneens op diverse manieren aangeboden waaronder tekst met illustraties of ondersteund met dia's, televisie of computer. Ter afsluiting en toetsing werd de PSE-test aangeboden die bestond uit een tekentoets, een identificatietoets, een terminologietoets en een begripstoets waarin ook werd gevarieerd in representatievormen. Dwyer (1994) trok naar aanleiding van het PSE-onderzoeksprogramma diverse algemene conclusies. Voor het hier beschreven onderzoek worden er daarvan kort enkele opgesomd, die vervolgens in de volgende hoofdstukken worden uitgediept.

1. De combinatie van visueel leren en visueel toetsen bevordert in veel onderwijssituaties de prestatie. Visueel toetsen zou over het algemeen een geschikte strategie zijn voor het toetsen van kennis die de lerende heeft opgedaan tijdens de visuele instructie.
2. Een toenemende mate van realisme in een visuele representatie leidt niet automatisch tot een corresponderende toename in informatieopname door een lerende.

Ook het overgrote deel van ander onderzoek over visuele kenmerken richt zich vooral op de rol die visuele representaties hebben in het leerproces als deze in combinatie met tekstueel instructiemateriaal worden aangeboden. Anglin, Vaez en Cunningham (2004) hebben getracht door literatuuronderzoek inzicht te krijgen in de rollen van statische en dynamische visuele representaties in het leerproces. Daartoe bestudeerden zij eerst deelprocessen in het leerproces in relatie tot perceptie en geheugenmodellen om daarna diverse onderzoeken naar kennisverwerving met statische of dynamische visuele representaties met elkaar

te vergelijken. Na bestudering en vergelijking van 90 onderzoeken, waaronder de studies van Dwyer met de PSE-methode, kwamen de auteurs tot een aantal conclusies waarvan er twee voor het hier beschreven onderzoek direct belang hebben. Ten eerste dat statische visuele representaties die tekstueel instructiemateriaal ondersteunen het leerproces kunnen faciliteren maar dat deze effecten zich niet consistent voordoen in alle leersituaties. Ten tweede dat realisme een relevante factor is in visuele representaties die nader onderzoek behoeft. Net als Dwyer concluderen Anglin, Vaez en Cunningham dat er een niet-lineaire relatie is tussen de mate van realisme in representaties en het leereffect dat het tot gevolg heeft. Nadere bestudering van realisme in het hier gepresenteerde onderzoek, meer specifiek hoofdstuk 2, zal uitwijzen dat de term 'realisme' niet eenduidig wordt gedefinieerd, wat het trekken van algemene conclusies bemoeilijkt. Zo wordt in veel onderzoeken op verschillende wijze gevarieerd in realisme door mate van detailinformatie, kleurcodering en beweging. Anglin, Vaez en Cunningham geven aan dat hun conclusies zijn getrokken op basis van de door hen vergeleken onderzoeken die voornamelijk waren gericht op de effecten van representaties als zodanig en veel minder op hun functie en rol tijdens het leerproces. Zij bepleiten om verder onderzoek te stellen op theorie over perceptie en geheugenprocessen om een beter beeld te krijgen van het functioneren van visuele representaties in leerprocessen. Voor het hier beschreven onderzoek betekent dit dat de vraag welke visuele kenmerken de te ontwerpen toets moet hebben om een goede afspiegeling te zijn van de met het realistische object verbonden leerstof enkel beantwoord kan worden als er inzicht is in de wijze en inrichting van de instructie samen met inzicht in de wijze waarop visuele informatie wordt opgenomen, opgeslagen en opgehaald. In hoofdstuk 2 wordt dit aspect verder uitgediept.

De centrale rol van het realistische object (de rat) in de instructie leidt als vanzelf tot het bestuderen van realisme als belangrijk kenmerk van visuele representaties. Een eerste oriëntatie in de literatuur laat inderdaad een wisselend beeld zien over het effect van realisme. In een onderzoek naar het effect van de mate van realisme op de herkenbaarheid van plaatjes was de vraag in hoeverre herkenning plaatsvond op basis van globale informatie zoals contouren in een tekening of op basis van informatieve details die aanwezig zijn in foto's (Loftus & Bell, 1975). De metingen werden viermaal herhaald en uit de resultaten bleek dat foto's beter werden herkend dan tekeningen. Volgens de onderzoekers werd dit veroorzaakt doordat foto's veel meer informatie aanbieden door de vele gedetailleerde visuele

kenmerken in vergelijking tot tekeningen. Uit soortgelijk onderzoek van Borges, Stepnowsky en Holt (1977) blijkt dat herkenning van kleurenfoto's significant hoger is dan de herkenning van zwart-wit foto's en herkenning van deze twee modaliteitvormen significant hoger dan herkenning van woorden. In tegenstelling tot deze resultaten, resulteert onderzoek van Nelson, Metzler en Reed (1974) juist in weinig significant verschil in herkenning tussen foto's en tekeningen, zelfs na zeven weken. Wel was er sprake van een plafondeffect in herkenning als de test direct na het experiment werd uitgevoerd. Het toevoegen van meer detail leidde vervolgens niet tot het beter herinneren van de plaatjes. Onderzoek van Ritchey (1982) die een vergelijking maakt tussen het herkennen van woorden, gedetailleerde tekeningen en contourtekeningen bij kinderen en volwassenen leverde weer andere resultaten op. Uit dit onderzoek blijkt dat contourtekeningen beter werden herkend dan gedetailleerde tekeningen en beide weer veel beter dan woorden.

De tegenstrijdige resultaten van deze onderzoeken laten zien dat de werking van de factor realisme geen eenduidig beeld geeft en kennelijk afhangt van de interactie met andere factoren die in de verschillende onderzoeken aan de orde waren. In hoofdstuk 2 wordt hier dieper op ingegaan om inzicht te krijgen in de mogelijke omstandigheden waaronder het aspect realisme in visuele representaties van het reële object adequaat kan bijdragen aan het toetsen van een construct. Dit heeft ertoe geleid dat realisme als variabele in het onderzoek naar de te ontwikkelen toets wordt meegenomen.

Naast realisme is er ook onderzoek gedaan naar de hoeveelheid contextuele informatie in visuele representaties en in hoeverre die van invloed is op de herkenning. Hiertoe is er onderzoek verricht naar het herkennen van gezichten (Tanaka & Farah, 1993; Moscovitch & Moscovitch, 2000; Cave & Kosslyn, 1993) en naar de wijze waarop herkenning plaatsvindt. Moscovitch en Moscovitch deden onderzoek naar de vraag of bij identificatie van gezichten de proefpersonen zouden letten op het gezicht als geheel (holistic-based) met contouren of op elementaire onderdelen (part-based) van het gezicht zoals neus en ogen zonder contouren. Uit de resultaten bleek dat gezichten met de juiste ruimtelijke relaties en in de correcte contouren het best werden herkend en dat de herkenning daalde naarmate het beeld meer werd verstoord waarbij herkenning van afzonderlijke onderdelen altijd hoog was. Deze bevindingen zouden ook gelden voor het herkennen van andere objecten. Palmer (1975) deed onderzoek naar de herkenning van objecten in (a) aanwezigheid van contextuele informatie, (b)

afwezigheid van contextuele informatie en (c) aanwezigheid van verkeerde contextuele informatie. Het ging daarbij om objecten binnen de context van een keuken. De objecten werden het best herkend in aanwezigheid van de correcte contextuele informatie en het slechtst werden objecten herkend in aanwezigheid van de verkeerde contextuele informatie. Te verwachten is dat voor het te toetsen construct in dit onderzoek vergelijkbare resultaten worden gevonden. Tijdens de instructie krijgen studenten de afzonderlijke anatomische onderdelen binnen de correcte context van biologische structuren te zien zoals het patroon waarin de onderdelen georganiseerd zijn met bijbehorende verhoudingen. Het is te verwachten dat het aanbieden van context beter aansluit op het toetsen van het construct dan het niet aanbieden daarvan. Voor het hier beschreven onderzoek is context als variabele meegenomen omdat het een essentieel visueel kenmerk is waarbij onderzocht wordt of het gevonden principe ook geldt voor de hier gepresenteerde toetsproblematiek. Het aanbieden van context levert voor de waarnemer vooral informatie op over het totaalbeeld van de anatomische structuur waarin de afzonderlijke objecten gezamenlijk één geheel vormen.

Samengevat zal in het hier beschreven onderzoek worden beargumenteerd dat voor de ontwikkeling van een toets met statische externe representaties van het bestudeerde object, de factoren realisme en context belangrijke visuele kenmerken zijn waarvan nog onvoldoende bekend is over hun bijdrage aan het toetsen van het geleerde construct. Om die reden zullen beide variabelen bij de ontwikkeling van de toets worden bestudeerd door middel van experimenten om inzicht te krijgen in hun bijdrage aan het te toetsen construct voor het hier gepresenteerde toetsprobleem. Mogelijk leveren de onderzoeksresultaten ook informatie op die zich leent voor generalisatie en die als zodoende bijdraagt aan de discussie over het gebruik van realisme en context in toetsmateriaal.

In hoofdstuk 2 wordt door literatuurstudie dieper ingegaan op deze materie.

De onderzoeksvraag

Om de eerder benoemde problemen aan te pakken, is kennis vereist van de precieze onderwijsdoelen – het aan te leren construct – in relatie tot visueel leren en visueel toetsen. Om het hoofd te bieden aan de eerder genoemde arbeidsintensieve evaluatie van de leerresultaten, de onzekerheid over de mate van objectiviteit in beoordeling en de late feedback op prestaties leidt dit tot de conclusie dat enkel het ontwikkelen van een toets die een afspiegeling is van de

leercondities onvoldoende is. Het onderzoek dient niet te starten bij het ontwikkelen van een toets maar bij een analyse van het leerproces dat plaatsvindt gelet op de combinatie van instructie en toetsing en de functies die representaties daarin vervullen. De eerder gepresenteerde conclusies van Anglin, Vaez en Cunningham (2004) onderschrijven dit. Relevante onderwerpen in de instructie zijn de effectiviteit van tekenen als instructiemiddel in het bereiken van de leerdoelen, de aansluiting daarvan op een adequate toetsmethodiek, de mogelijkheden voor snelle feedback. De vraag naar een adequate toets voor het toetsen van kennis opgedaan door observatie van concreet materiaal wordt met het bovenstaande uitgebreid met de vraag of de instructie ook adequaat is voor het bereiken van de leerdoelen. Het gevolg is een verbreding van het onderzoekstraject van het oplossen van een toetsproblematiek naar het bestuderen van de onderwijssituatie als geheel uitgaand van een samenspel tussen instructie en toets. De beoogde onderwijsvernieuwing beperkt zich daarmee niet alleen tot het ontwikkelen van een toets maar richt zich op de gehele onderwijssituatie.

Dit heeft geleid tot onderstaande algemene onderzoeksvraag

Wat is een effectieve en efficiënte combinatie van instructie en toetsing voor het bereiken van onderwijsdoelstellingen die verbonden zijn met leren door observeren van concreet biologisch materiaal.

Effectiviteit en efficiëntie zoals hier bedoeld worden het best omschreven door Molenda (2009). Effectiviteit heeft te maken met de mate waarin de leerdoelen worden behaald. Efficiency in onderwijskundige context behelst het ontwerpen, ontwikkelen en bewerkstelligen van instructie op dusdanige wijze dat de beoogde resultaten worden bereikt tegen zo weinig mogelijk kosten. Beide factoren zijn niet afzonderlijk van elkaar te bezien. Efficiënte instructie is zinloos als de leerdoelen niet worden behaald. Effectieve instructie die buitensporige middelen gebruikt of niet de juiste mensen bereikt is onvoldoende productief. Molenda koppelt efficiëntie en effectiviteit aan productiviteit. Een onderwijssituatie is productief als deze zowel effectief als efficiënt is. Vertaald naar het hier beschreven onderzoek is effectiviteit het behalen van de gewenste onderwijsdoelstellingen terwijl de daaraan gekoppelde toets het behalen van die doelen correct meet. Ten opzichte van de bestaande situatie kan de efficiëntie worden verbeterd door het reduceren van de arbeidsintensieve evaluatie van de leerresultaten. Voor het oplossen van de andere twee problemen mag een

verbetering van zowel de effectiviteit als de efficiëntie worden verwacht door het ontwikkelen van een objectief beoordeelbare toets die tevens snelle en adequate feedback mogelijk maakt. Door de relatie te leggen tussen *instructie en toetsing* in de algemene vraag wordt aangegeven dat voor een optimale productiviteit het gehele proces van instructie en toetsing van belang is. Dit resulteert voor het concrete probleem dat in dit onderzoek wordt aangepakt in een onderzoekstraject waarin wordt begonnen met het bestuderen van het leerproces in de bestaande instructie waarin getekend wordt om van daaruit te komen tot criteria voor de ontwikkeling van een adequate combinatie van instructie en toets voor het oplossen van het toetsprobleem. Het laatste onderdeel van de algemene vraag dat zich richt op het bereiken van de *onderwijsdoelstellingen die verbonden zijn met leren door observeren van concreet biologisch materiaal* impliceert de eis dat de te ontwikkelen toets een afspiegeling moet zijn van het te leren construct.

Om de algemene onderzoeksvraag te kunnen beantwoorden is deze vertaald naar specifiekere onderzoeksvragen en hypothesen die hebben geresulteerd in drie deelonderzoeken. Deze betreffen een ontwerpgericht onderzoek voor het ontwikkelen van een geschikte instructie-toets-combinatie als alternatief voor de bestaande situatie, het ontwerpen en testen van het bijbehorende toetsconcept en tot slot het uitvoeren van een tweetal experimenten voor het kiezen van de definitieve vormgeving van de toets in relatie tot de onderzoeksvariabelen realisme en context.

In het eerste experiment in het hier gepresenteerde onderzoek wordt gezocht naar een voor dit project geschikte instructie-toets-combinatie (hoofdstuk 3). De bestaande situatie wordt geanalyseerd waarbij de evaluatie van de instructie vooral gericht is op bestudering van de leerdoelen van de instructie en de rol van het tekenen daarin. Beargumenteerd zal worden dat de leerdoelen van de instructie gericht zijn op de statische configuratie van een biologisch systeem (Hagerty, 2005), meer specifiek het kunnen herkennen en benoemen van anatomische structuren en het kunnen aangeven van de ruimtelijke relaties tussen deze structuren rekeninghoudend met de relatieve grootte van elke structuur in relatie tot andere structuren. Deze leerdoelen worden in de bestaande instructie bereikt door de student nauwkeurig te laten observeren, het geobserveerde te laten tekenen en de getekende onderdelen te laten benoemen. Literatuuronderzoek levert op dat het tekenen inderdaad een effectieve instructiemethode is (Lansing, 1984). Reden daartoe is dat op deze wijze

doelgericht aandacht wordt besteed aan die visuele kenmerken die van belang zijn te midden van een veelheid aan visuele informatie (Levie, 1987). Uit de literatuur blijkt ook het aanbieden van een standaardtekening met een benoemingsopdracht een effectieve instructiemethode te zijn voor het richten van de aandacht op relevante visuele kenmerken vergelijkbaar met tekenen als instructiemethodiek (Dean en Kulhavy, 1981). In hoofdstuk 2 en 3 wordt dieper in gegaan op deze onderzoeken die in combinatie met een analyse van de oorspronkelijke instructie leiden tot het voorstel om de oorspronkelijke instructie zo te vernieuwen dat niet tekenen als instructiemethodiek wordt uitgevoerd maar dat in plaats daarvan een standaardtekening wordt aangeboden met de opdracht om bij de bestudering van de diverse dissectielagen van de rat de relevante objectonderdelen te lokaliseren en te benoemen. Dit betekent dat er in de nieuwe situatie geen zelfgemaakte tekeningen meer zijn die de basis kunnen vormen voor de terugkoppeling en de beoordeling van het geleerde. In hoofdstuk 3 wordt beschreven dat voor de terugkoppeling tijdens de instructie een groepsgewijs mondeling terugkoppelingsmoment is ingelast en voor de beoordeling een objectief toetsbare potlood-en-papiertoets is ontwikkeld. Deze toets behoort samen met de vernieuwde instructie een geschikte instructie-toets-combinatie te zijn om de studenten de leerdoelen van het dissectiepracticum te laten bereiken.

De vernieuwde instructie wordt geëvalueerd in een experiment waarin beide instructievormen (de bestaande instructie en de vernieuwde instructie) met elkaar worden vergeleken.

Hierbij gaat het om de volgende onderzoeksvraag:

Is de vernieuwde instructie gelijkwaardig aan de bestaande instructie voor het bereiken van de onderwijsdoelen van het dissectiepracticum anatomie van de rat?

In hoofdstuk 3 worden de onderzoeksresultaten gepresenteerd alsook de conclusies met als eindresultaat dat de vernieuwde instructie verder als uitgangspunt kan worden gebruikt voor het ontwikkelen van een oplossing van het toetsprobleem dat de aanleiding vormt voor dit onderzoeksproject.

Het onderzoek wordt vervolgd met het ontwerpen en testen van het bijbehorende toetsconcept. De objectief toetsbare potlood-en-papiertoets die in het kader van de ontwikkeling van de vernieuwde instructie is ontworpen, heeft een voorlopig karakter met betrekking tot twee aspecten. Het eerste aspect betreft de effectiviteit

van de toets. De inhoud die in de toets is gerepresenteerd wordt beschouwd als een bruikbare maar eerste benadering van het te toetsen construct. In de volgende fase van het onderzoek moet die inhoud worden verfijnd. Het tweede aspect betreft de efficiëntie van de toets. Op dit punt wordt in de verdere ontwikkeling gekozen voor de constructie van een gecomputeriseerde toets met computer-based test items (CBT-toets). Bij de grote aantallen studenten waarom het gaat en de stabiele leerstof die getoetst moeten worden, mogen van CBT-toetsen duidelijke efficiëntievoordelen worden verwacht bij geautomatiseerde toetsafname. Naar aanleiding van de resultaten uit hoofdstuk 3 zal blijken dat, vooraf aan de constructie van de toets, eerst een evaluatie van het te leren construct noodzakelijk is. Met docenten en vakinhoudelijke experts is dit opgepakt en de resultaten daaruit zijn gebruikt voor het bijstellen van de definitie van het te leren construct. Pas daarna kan worden gestart met de doorontwikkeling van de toets naar een CBT-toets. Dat gebeurt volgens het stappenplan voor testconstructie van Oosterveld en Vorst (1998), dat waar nodig aangepast is aan de huidige onderzoekssituatie. Het eindresultaat is een toets die bestaat uit drie toetstaken te weten een herkenningstaak, een benoemingstaak en tot slot een taak gericht op het leggen van functionele relaties tussen anatomische structuren. De onderbouwing hiervan is onderdeel van hoofdstuk 4.

Na de constructie van de CBT-toets wordt het traject vervolgd met experimenten die gericht zijn op twee onderzoekslijnen namelijk in hoeverre de toets inderdaad effectief is in het benaderen van het geleerde construct en als zodanig met de aangepaste instructie een geschikte instructie-toetscombinatie vormt, en wat de beste vormgeving is van de toetsitems in de CBT-toets met betrekking tot de variabelen realisme en context. Het onderzoek is nodig omdat zoals eerder al is besproken de variabelen realisme en context volgens de literatuur weliswaar belangrijke variabelen zijn voor het in een toets representeren van een construct, maar dat niet duidelijk is wat ze bijdragen aan het toetsen van het construct voor het hier gepresenteerde toetsprobleem. Vertaling van elke variabele naar een concrete vorm resulteerde in twee uiterste representaties. Zo is realisme geoperationaliseerd in een fotovorm die de realistische weergave van de anatomische objecten vrijwel benadert en in een schematische vorm die enkel de vormen van de objecten weergeeft. De variabele context is geoperationaliseerd door contextuele informatie aan te bieden in de vorm van de contouren van de rat waar tegenover visualiseringen zonder context staan waarbij de contouren van de rat zijn weggelaten. Voor het onderzoek heeft dit geleid tot vier toetsvormen

(foto's zonder context, schema's zonder context, foto's met context, schema's met context) die worden gebruikt in twee experimenten voor het beantwoorden van de volgende specifieke onderzoeksvragen:

Meet een toets met schema's hetzelfde visuele construct als een toets met foto's?

Meet een toets met context hetzelfde visuele construct als een toets zonder context?

Het doel van het onderzoek rond de variabelen realisme en context is het ontwikkelen van een zo goed mogelijke toets voor het onderhavige toetsprobleem. Om die reden is het onderzoek uitgevoerd in twee ronden. Het eerste experiment dat is uitgevoerd had naast het toetsen van de hypothesen ook tot doel het eerste toetsontwerp te evalueren en waar mogelijk te verfijnen. Na verfijning is het experiment herhaald in een tweede experiment op basis waarvan het ontwerp van de instructie-toets-combinatie verder kon worden verfijnd. De resultaten geven uitsluitsel over de geformuleerde hypothesen als component van de beantwoording van de algemene onderzoeksvraag van het onderzoek. In de hoofdstukken 5 en 6 worden de experimenten gepresenteerd.

Beantwoorden van de onderzoeksvraag vanuit een ontwerpgerichte onderzoeksmethodologie

Het hier gepresenteerde onderzoek is op de eerste plaats een onderzoek naar een te ontwerpen interventie voor een toetsprobleem binnen een bestaande onderwijssetting: het dissectiepracticum anatomie van de rat. Het beschreven toetsprobleem is echter algemener van aard dan het huidige practicum binnen de biologie dat de aanleiding is voor het onderzoek. Het algemene probleem betreft het toetsen van kennis die wordt ontleend aan realistische objecten. Beargumenteerd is dat voor het ontwikkelen van een adequate toets het essentieel is om de gehele onderwijssituatie te bestuderen voor een nauwe afstemming tussen instructie en toetsing gelet op effectiviteit en efficiëntie.

Om het onderzoeksdoel te bereiken is gekozen voor een ontwerpgerichte onderzoeksaanpak. Centraal in ontwerpgericht onderzoek is het op systematische wijze ontwerpen, ontwikkelen en evalueren van een educatieve interventie binnen de reële onderwijssetting (Van den Akker, Gravemeijer, Mc Kenney & Nieveen, 2006). Tegelijkertijd is ontwerpgericht onderzoek gericht op theorievorming over de bijdrage van de specifieke interventie aan theorieën over leren en onderwijzen. Het is daarmee zowel pragmatisch als theoretisch van aard en beantwoordt aan de onderzoeksdoelstellingen van het hier gepresenteerde

onderzoek. Door het toepassen van een ontwerpgerichte benadering komen zowel de toets als de vernieuwde instructie pragmatisch tot stand. Tegelijkertijd wordt getracht om bij het ontwikkelen van de toets algemene richtlijnen te ontleen voor de toetsmethodiek die bijdragen aan de theorievorming over toetsen met visuele representaties.

Van den Akker, Gravemeijer, Mc Kenney en Nieveen (2006) geven naast het bovenstaande nog een kenmerk van ontwerpgericht onderzoek dat van toepassing is op het hier gepresenteerde onderzoek. Door het actief betrekken van inhoudelijke experts en gebruikers in het onderzoek is het mogelijk om direct te interveniëren binnen de context waarin het onderzoek zich afspeelt. Hun inbreng kan bijdragen tot verbetering van de onderwijssetting alsook het ontwerpen van de interventie. Door gebruikers binnen de reële context te betrekken bij de evaluatie van de ontworpen interventie wordt ook informatie gegenereerd over de toepasbaarheid daarvan. Dit geeft informatie over de effectiviteit en efficiëntie van de ontworpen interventie, de toets, in deze specifieke onderwijssituatie.

In het hier gepresenteerde onderzoek zijn inhoudelijk experts actief betrokken bij het onderzoek als onderdeel van het onderzoeksteam. Deze inhoudelijk experts droegen bij aan de analyse van het kennisgebied en de onderzoeksproblematiek, aan de evaluatie van de instructie, de formulering van de leerdoelen en het te leren construct, de constructie en de evaluatie van de toets. Met studenten als gebruikers is de toets geëvalueerd door hen tijdens de toets retrospectief schriftelijk weer te laten geven over de wijze waarop zij elk vraag beantwoordden.

Mc Kenney, Nieveen en Van den Akker (2006) wijzen op enkele dilemma's waar een onderzoeker in ontwerpgericht onderzoek tegenaan loopt. Allereerst de dubbele rol die de onderzoeker heeft als uitvoerder van het onderzoek en als beoordelaar van het doorlopen onderzoek. Door te werken in de praktijk ontstaat een meestal langdurig samenwerkingsverband met de belanghebbenden zoals docenten, studenten en opdrachtgevers. De onderzoeker is ontwerper (uitvoerder) en als zodanig een voorstander van het te doorlopen onderzoek. Tegelijkertijd dient deze ook kritisch te zijn en op objectieve manier het doorlopen onderzoek te evalueren. Dit kan rolverwarrend zijn voor deelnemers aan het onderzoek en voor de onderzoeker. Plomp (2009) adviseert hierin dat de onderzoeker (1) open staat voor professionele verificatie en kritiek van mensen buiten het project, (2) de vuistregel toepast om aan de start van het onderzoek in de rol te zijn van de

dominante of creatieve onderzoeker en deze gaandeweg bij te stellen naar de rol van de kritische onderzoeker en (3) zich inspant voor kwalitatief goed onderzoek dat zich kenmerkt door consistentie, triangulatie, empirisch testen van zowel bruikbaarheid als effectiviteit van de interventie, systematische documentatie alsook aandacht voor validiteit en betrouwbaarheid van data en instrumenten. In het hier gepresenteerde onderzoek is het gehele onderzoekstraject doorlopen binnen de Biologie waarbij het onderzoeksteam bestond uit de onderzoeker en de inhoudsdeskundigen op het gebied van de biologie. Door regelmatig overleg met enerzijds de inhoudsdeskundigen en anderzijds professionele verificatie met externe onderwijskundige onderzoekers is in het hier gepresenteerde onderzoek getracht de rolverdeling (onderzoeker vs. betrokkene) in balans te houden en uit te voeren. Daarnaast is het gehele onderzoek empirisch uitgevoerd en geëvalueerd, systematisch gedocumenteerd en zijn er ontwerpcycli doorlopen. Overigens zijn alle activiteiten binnen de ontwerpcycli empirisch getest.

Het volgende dilemma betreft het aanpassingsvermogen van het onderzoek en de onderzoeker. Mc Kenney et al (2006) wijzen erop dat gegeven de dynamiek van de reële onderwijssetting het essentieel is om waarnodig het onderzoek te kunnen aanpassen. Elke onderzoekscyclus dient volgend te zijn op de resultaten afkomstig uit eerdere. Tegelijkertijd is een onderzoek dat continue verandert ook een teken van zwakte. In het hier gepresenteerde onderzoek is continue elk deelonderzoek steeds geplaatst binnen het overzicht van de gehele onderwijssituatie. Aanpassingen vonden enkel plaats als deze voortvloeiden uit eerdere onderzoeksresultaten of analyses.

Het gehele onderzoek is uitgevoerd met in achtneming van bovenstaande dilemma's en getracht is daar een balans in te vinden variërend van participerend onderzoeker bij de instructie-innovatie en de ontwikkeling van de toets, tot objectief onderzoeker van de toetseigenschappen voor het genereren van data met twee doelen: conclusies trekken over het te leren en te toetsen construct en de te ontwerpen toets en op basis daarvan besluiten nemen over de toetsvorm voor het onderhavige onderwijs.

Totaaloverzicht in het onderzoek

In het hier beschreven hoofdstuk is een overzicht gegeven van de onderzoeksproblematiek. De eerste analyse daarvan heeft geleid tot de formulering van de algemene onderzoeksvraag.

Wat is een effectieve en efficiënte combinatie van instructie en toetsing voor het bereiken van onderwijsdoelstellingen die verbonden zijn met leren door observeren van concreet biologisch materiaal?

In het onderzoeksproces van ontwerpgericht onderzoek wordt volgens Plomp (2009) een drietal fasen doorlopen. Het vooronderzoek waarin een inhoudelijke analyse en een behoefteanalyse plaatsvindt alsook een literatuuronderzoek dat resulteert in een conceptueel of theoretisch kader van het onderzoek. Daarna volgt de prototypefase waarin de beoogde interventie wordt ontwikkeld. In het hier beschreven onderzoek worden in deze fase twee afzonderlijke onderzoekslijnen uitgevoerd die gezamenlijk bijdragen aan het beantwoorden van de onderzoeksvraag. De voornaamste onderzoekslijn is de ontwikkeling van een instructie-toetscombinatie die dekkend is voor het construct. De tweede onderzoekslijn is onderzoek naar de meest geschikte representatievorm voor het toetsen van dat construct. De derde en laatste fase is de evaluatiefase die zich kenmerkt door summatieve evaluatie, in dit onderzoek ter beoordeling of de instructie-toetscombinatie ook voldoet aan de eerder opgestelde eisen. Effectiviteit van de interventie is hierin van belang. Deze fase kan ook resulteren in aanbevelingen voor de verbetering van de interventie. Gedurende alle activiteiten binnen deze fasen wordt systematisch gedocumenteerd en gereflecteerd. Deze fasering loopt als rode draad door het hier gepresenteerde onderzoek. De hoofdstukken 1 en 2 zijn een weergave van het vooronderzoek. Hoofdstukken 3, 4, 5 en 6 gaan in op de prototypefase en rapporteren over de twee onderzoekslijnen die weliswaar afzonderlijk worden uitgevoerd maar gezamenlijk bijdragen aan het beantwoorden van de onderzoeksvraag in het betreffende hoofdstuk. Zo wordt in hoofdstuk 3 de bestaande instructie in een experiment geëvalueerd maar wordt er ook een eerste toets ontwikkeld in een pen-en-papier-versie die de effectiviteit van de instructie toetst. In Hoofdstuk 4 wordt, naar aanleiding van de resultaten uit hoofdstuk 3, het te leren construct geëvalueerd met de docenten. Daarnaast wordt ook de CBT-toets ontwikkeld op basis van het vooronderzoek, de bevindingen met de papierentoets en evaluatieresultaten van het te leren construct naar aanleiding van interviews. In de hoofdstukken 5 en 6 wordt de kwaliteit van de instructie-toetscombinatie onderzocht, waarnodig bijgesteld en wordt getracht de specifieke onderzoeksvragen te beantwoorden met betrekking tot de variabelen realisme en context. De summatieve evaluatie is vastgelegd in hoofdstuk 7 dat beschrijft in hoeverre de onderzoeksvraag is beantwoord ten aanzien van een effectieve en

efficiënte instructie-toets-combinatie voor het bereiken van de onderwijsdoelstellingen. Daarbij wordt ingegaan op de verkregen resultaten uit de prototypefase voor de beantwoording van de specifieke onderzoeksvragen.

In figuur 1.2 is het onderzoekstraject schematisch weergegeven waarbij tevens is aangegeven hoe de verschillende onderdelen over de hoofdstukken zijn verdeeld.

Figuur 1.2 Overzicht onderzoekstraject

In overeenstemming met figuur 1.2 wordt getracht de algemene onderzoeksvraag in de komende hoofdstukken te beantwoorden voor het particuliere geval van het dissectiepracticum anatomie van de rat door het ontwikkelen van een met onderzoek onderbouwde instructie-toetscombinatie. Beoogd is dat deze instructie-toetscombinatie op basis van het begeleidende onderzoek metterdaad als een antwoord op de algemene vraag kan worden beschouwd. In het afsluitende hoofdstuk 7 wordt de balans opgemaakt voor wat betreft het succes van de ontwikkelde interventie en de toepasbaarheid in vergelijkbare onderwijssituaties, alsmede de bijdrage aan de kennis over visueel leren (de instructie) en visueel toetsen (de toets) met betrekking tot de variabelen realisme en context.

HOOFDSTUK 2

Visueel leren en toetsen

In het voorgaande hoofdstuk is de algemene onderzoeksvraag gepresenteerd waarbij is gesteld dat alleen het ontwikkelen van een toets die een afspiegeling is van de leercondities onvoldoende is voor het beantwoorden van deze onderzoeksvraag. Gelet op de combinatie van instructie en toetsing, en de functies die representaties daarin vervullen is beargumenteerd dat het onderzoek dient te starten met een analyse van het leerproces dat plaatsvindt. Doel van het hier gepresenteerde hoofdstuk is deze analyse uit te voeren en meer inzicht te krijgen in de wijze waarop leerprocessen en daarbij betrokken representaties van invloed kunnen zijn op instructie en toetsing. Het eindresultaat van dit hoofdstuk is een theoretische basis die een aantal argumenten oplevert om (a) instructie en toets in te richten voor het bereiken van de onderwijsdoelstellingen die een weergave zijn van het te leren construct, (b) de geformuleerde onderzoeksvraag te vertalen naar specifiekere onderzoeksvragen en (c) variabelen te definiëren voor de experimenten ter beantwoording van die specifiekere onderzoeksvragen.

2.1 TOELICHTING OP DE VRAAGSTELLING EN DE AANPAK VAN BESCHIKBARE LITERATUUR

De eerder gepresenteerde algemene onderzoeksvraag luidt als volgt:

‘Wat is een effectieve en efficiënte combinatie van instructie en toetsing voor het bereiken van onderwijsdoelstellingen die verbonden zijn met leren door observeren van concreet biologisch materiaal’

Deze vraagstelling is opgesteld vanuit een aantal overwegingen zoals beargumenteerd in hoofdstuk 1. Kort weergegeven zijn het de volgende:

- De prestatie op de toets dient uit te drukken of de persoon in kwestie de veronderstelde kennis over een bepaald domein heeft opgedaan;

- Deze kennis over een bepaald domein wordt beschouwd als een veronderstelde eigenschap die de te toetsen persoon al of niet beheerst. Voor de op deze manier gedefinieerde kennis wordt de term “construct” gebruikt;
- Een goede constructdefinitie waarin deze veronderstelde eigenschap wordt beschreven is een belangrijk uitgangspunt voor de constructie van een kwalitatief goede toets die er daarmee een afspiegeling van is;
- Het combineren van visueel toetsen en visueel leren bevordert in veel onderwijssituaties de prestatie en daarmee het bereiken van de onderwijsdoelstellingen;
- Bovenstaande argumenten impliceren dat de onderwijsdoelstellingen moeten beantwoorden aan het te leren construct en daar een afspiegeling van moeten zijn.

Het hier gepresenteerde onderzoek is op basis van deze argumenten verbreed van het oplossen van een toetsprobleem naar het bestuderen van de onderwijssituatie als geheel uitgaand van een samenspel tussen instructie en toets. Dat heeft in dit hoofdstuk geleid tot een analyse van het leerproces dat plaatsvindt gelet op de combinatie instructie en toetsing en de functie die representaties daarin vervullen.

Alvorens dieper in te gaan op deze analyse enkele punten vooraf:

1. Onderzoek naar de bijdrage van visuele representaties in instructies is vooral gericht op de vraag of teksten met visuele representaties een meerwaarde hebben in het leerproces boven teksten zonder. Dit blijkt bij bestudering van de beschikbare literatuur over leren met representaties (visueel leren) en geldt ook voor onderzoeksvragen ten aanzien van toetsing met representaties (visueel toetsen). De analyse zoals gepresenteerd in dit hoofdstuk maakt vooral gebruik van conclusies uit dit type onderzoek waarbij wordt toegespitst op het vergelijken van verschillende representatievormen en de betekenis daarvan voor instructie en toetsing.
2. Het reële object is het studieobject dat in de onderwijssituatie van dit onderzoek wordt aangeboden aan de lerende. Het leerproces is er op gericht dat de lerende zelf de aangeboden informatie over de kenmerken van dat object vertaalt naar een afbeelding (visuele representatie). Voor het toetsen wordt het reële studieobject vervangen door afbeeldingen en wordt van de lerende verwacht dat deze de objectkenmerken die zijn weergegeven in de representatie, correct kan koppelen aan dan wel terug redeneren naar het reële object. Bij bestudering van de beschikbare literatuur blijkt het steeds te gaan om onderzoek naar de effecten van visuele representaties van reële objecten

zonder dat die reële objecten deel uit maken van de onderzoekssituatie (zie ook het vorige punt). Dat betekent dat de hier besproken literatuur over visuele representaties handelt over onderzoekssituaties waarin van de lerende wordt verwacht dat deze zowel tijdens het leerproces als tijdens het toetsproces telkens de vertaalslag maakt van de representatie van een object naar het reële object zonder de directe aanwezigheid van dat object. Dit wijkt af van de hier gepresenteerde onderzoekssituatie waarbij het reële object een centrale rol speelt in de instructie. Literatuur over onderzoek dat één op één vergelijkbaar is met de hier gepresenteerde onderzoekssituatie was tot op het moment van het uitgevoerde onderzoek niet beschikbaar. Verondersteld wordt dat de mentale processen die in de hier gepresenteerde onderzoekssituatie aan de orde zijn toch voldoende overeenkomst vertonen met de in dit hoofdstuk besproken literatuur om voor het onderzoek relevante conclusies te kunnen trekken.

3. Het te bestuderen reële object in de instructie van het hier gepresenteerde onderzoek is concreet biologisch materiaal in de vorm van de rat. De instructie heeft de vorm van een dissectiepracticum. De onderwijsdoelstellingen van het huidige practicum zijn: (a) het herkennen van anatomische structuren van de rat aan vorm, kleur en textuur, (b) het kunnen benoemen van deze anatomische structuren met de Latijnse namen en (c) het kunnen aangeven van de ruimtelijke relaties tussen deze structuren waarbij rekening wordt gehouden met de relatieve grootte van elke structuur gerelateerd aan andere structuren. Deze doelstellingen vormen een afspiegeling van het te leren construct.
4. De centrale rol van het reële object in de instructie indiceert het bestuderen van realisme als belangrijk kenmerk in de relatie tussen het reële object en representaties daarvan. Het reële object bevat als zodanig de meeste informatie over het uiterlijk van dat object in vergelijking tot elke representatie daarvan en is daarmee de meest realistische versie. Daarmee bevat elke visuele representatie van het object, in vergelijking tot het object zelf, een lager realiteitsgehalte. In de Van Dale (1999) is de definitie voor realisme: een plaatje dat realistisch is als deze een kopie is van de visuele kenmerken van de referent (het gevisualiseerde object). Omdat het weergeven of representeren van de werkelijkheid van de visuele objectkenmerken erg divers is en vele varianten kent, is voor het doel van dit hoofdstuk een iets ruimere definitie op zijn plaats: een visuele representatie van een object is realistisch in de mate waarin de werkelijkheid van de visuele kenmerken van dat object zijn

weergegeven. De mogelijke diversiteit aan representatievormen doet daarbij, in het hier gepresenteerde onderzoek, veronderstellen dat er sprake is van een rangschikking van meer naar minder realistische objectweergaven in een samenhangende reeks van varianten in de weergave van de werkelijkheid. Het hoogst in deze reeks is het reële object en elke weergave van dat object is lager in de reeks doordat deze een in visueel opzicht gereduceerde versie van de werkelijkheid weergeeft. Deze assumptie is gangbaar in veel onderzoek naar bijdragen van visuele representaties. De variabele realisme wordt in vele onderzoeken gevarieerd in de mate waarin een visuele representatie de werkelijkheid van een object weergeeft. Dit gebeurt voornamelijk door het manipuleren van de hoeveelheid detailinformatie. Foto's en tekeningen, al of niet met toevoegingen zoals kleurcoderingen, zijn bekende varianten.

Er zijn ook andere varianten mogelijk zoals het bieden van een compleet overzicht van het reële object in de vorm van (deel)objecten die al of niet worden weergegeven, al dan niet in combinatie met omliggende objecten rondom het reële object. In het hier gepresenteerde onderzoek kan daarbij worden gedacht aan de topografische ligging van diverse anatomische objecten in de rat die afzonderlijk dan wel in relatie tot andere daarnaast gelegen objecten kunnen worden getoond zoals de contouren van de rat.

In de verdere uitwerking van de literatuur in dit hoofdstuk wordt dit onderwerp verder uitgediept in relatie tot leren en toetsen.

De literatuur in dit hoofdstuk start met het scheppen van een algemeen kader waarna steeds meer wordt ingezoomd op de aandachtsgebieden die in dit onderzoek een belangrijke rol spelen. Daartoe worden eerst de cognitieve processen beschreven in termen van de keten perceptie, informatieverwerking en geheugenopslag. Vervolgens wordt ingegaan op visuele variabelen die daarbij een rol spelen, zoals vorm en functie. Met deze informatie wordt bekeken wat dat betekent voor instructies met representaties en er wordt ingegaan op de toetsvormen die de resultaten van visueel leren kunnen toetsen. Deze blijken divers waarbij nog weinig onderzoek voorradig is over de rol die representaties daarin spelen. De theoretische basis gepresenteerd in dit hoofdstuk vormt de basis voor nader onderzoek in de daarna volgende hoofdstukken.

2.2 COGNITIEVE PROCESSEN EN DE ROL VAN BETROKKEN REPRESENTATIES

In de literatuur geven analyses van diverse auteurs (Anglin, Vaez & Cunningham, 2004; Braden, 1996; Levie, 1987) naar de effecten van visuele representaties in leerprocessen aan dat er weinig samenhang is tussen de verschillende onderzoeken. Het genereren van algemeen geldende conclusies wordt daardoor bemoeilijkt. Levie (1987) geeft aan dat er in de literatuur ogenschijnlijk verschillende thematische eilanden bestaan waartussen weinig relaties worden gelegd. Onderzoekers naar geheugenmodellen maken bijvoorbeeld weinig gebruik van perceptietheorieën over representaties en vice versa. Zijn verklaring is dat theorievorming versplinterd plaatsvindt doordat onderzoekers opereren vanuit afzonderlijke onderzoeksgebieden zoals ontwikkelingspsychologie, perceptie, communicatie, kunst of onderwijspsychologie. Daarbij concentreren ze zich vooral op literatuur uit hun eigen onderzoeksgebied. In hun onderzoek richten ze zich op specifieke processen zoals aandacht of korte-termijn-geheugen waarin ze meer ingaan op de door stimuli opgewekte mentale processen dan op de kenmerken van de stimuli die deze mentale processen veroorzaken. Levie bepleit de integratie van de diverse onderzoeksgebieden ten aanzien van visuele representaties in de veronderstelling dat integratie wellicht nieuwe inzichten oplevert en nieuwe onderzoeksvragen. Een poging daartoe vindt plaats op het gebied van visuele geletterdheid ("visual literacy") (Braden, 1996; Hortin, 1994) waar vanuit uiteenlopende kennisgebieden de resultaten van onderzoek naar effecten van visuele representaties worden geanalyseerd en zo mogelijk samengevoegd ten behoeve van de theorievorming. In overeenstemming met het hier gepresenteerde onderzoek, richt visuele geletterdheid zich in de analyse op de cognitieve processen waarin visuele representaties een rol spelen. Het visual-literacy-onderzoek heeft bijgedragen aan het categoriseren van visuele representaties (Braden, 1996) waarbij is gebleken dat mensen niet alleen leren met visuele representaties (Anglin, Vaez en Cunningham, 2004; Braden, 1996; Levie, 1987) maar zich in hun denken ook daadwerkelijk visuele beelden vormen (Anglin, Vaez en Cunningham, 2004; Braden, 1996; Levie, 1987; Dwyer, 1978). Als zodanig kunnen ze vervolgens ook getoetst worden met visuele representaties (Dwyer en De Melo, 1983). Tegelijkertijd wordt geconcludeerd dat in tot dan toe verricht onderzoek onderbelicht is wat de functionaliteit is van visuele representaties in het leerproces (Anglin, Vaez en Cunningham, 2004; Braden, 1996; Levie, 1987). Visual literacy heeft niettemin de 'thematische eilanden' (Levie, 1987) voldoende bij elkaar kunnen brengen om een algemeen overzicht te geven

van het leren, denken en communiceren met representaties van waaruit ondermeer het bovenstaande kon worden geconcludeerd.

Deze aanpak wordt ook in het hier gepresenteerde onderzoek gehanteerd door vanuit verschillende onderzoeksgebieden de problematiek inzichtelijk te krijgen en richting te geven aan een adequate oplossing daarvan. Om dat te bereiken is er een selectie gemaakt van literatuur die hieronder wordt besproken voor het trekken van voor het hier gepresenteerde onderzoek relevante conclusies.

Cognitieve processen

Perceptietheorieën

Aan de basis van diverse onderzoeken naar herkenning van en omgang met beeldmateriaal liggen enkele theorieën (Gibson, 1972; Gombrich, 1969; Arnheim, 1974; Knowlton, 1966) ten grondslag over perceptie en interpretatie van visuele representaties zoals plaatjes. Voor het hier gepresenteerde onderzoek worden drie theorieën gepresenteerd die verschillende verklaringen geven over de wijze waarop informatie wordt opgenomen en de bijdrage aan de interpretatie daarvan.

Volgens Gibson (1972) wordt de interpretatie van een plaatje vooral bepaald door de stimulus vanuit directe perceptie. De waarnemer richt zich op de mate van overeenkomst tussen de inhoud van het plaatje, de representatie, enerzijds en anderzijds het reële object waarnaar het plaatje refereert. Herkenning van dat object in een plaatje vindt plaats doordat in de hersenen met elkaar corresponderende optische arrangementen, van het reële object en het bijbehorende representatieve plaatje, met elkaar worden vergeleken.

De constructivist Gombrich (1969) gaat een stap verder dan enkel het vergelijken van corresponderende optische arrangementen en benadrukt de rol van de waarnemer. De waarnemer zou, na het mentaal vergelijken van reëel object en representatief plaatje, de betekenis van het plaatje construeren vanuit eigen regels en verwachtingen. De waarnemer heeft daarbij een actieve rol in het bepalen welke visuele kenmerken uit een plaatje relevant zijn voor interpretatie met als gevolg dat de beeldopname voor hetzelfde object per individu kan verschillen.

De derde benadering is van Arnheim (1974) die beweert dat perceptie en interpretatie van plaatjes plaatsvindt vanuit de zogenaamde Gestalt-regels. Een waarnemer bekijkt een plaatje als geheel, past de Gestaltregels automatisch toe

wat bepalend is voor de interpretatie van dat plaatje. De Gestaltregels zijn regels die alle mogelijke 'unieke' relaties tussen kenmerken in een plaatje herleiden tot enkele basisprincipes over de organisatie van die kenmerken in vormen en patronen. Elk mens bezit het aangeboren vermogen voor het automatisch toepassen van deze regels. Arnheim (1986) wijst er vervolgens op dat perceptie niet gebaseerd is op passieve opname maar op het begrijpen van deze definieerbare vormen en patronen.

Deze drie theorieën zijn elk op hun eigen manier bepalend voor de wijze waarop een aantal kenmerken in visuele representaties wordt bestudeerd, waarbij realisme nogmaals als belangrijke factor wordt aangemerkt met elk een eigen definitie. Volgens Gibson is realisme extern bepaald en een functie van de mate van overeenkomst tussen de twee optische arrangementen van de representatie, zijnde het plaatje, en het reële object. Volgens constructivisten zoals Gombrich (1969) zijn de criteria voor mate van realiteit niet extern bepaald maar vooral intern door de waarnemer vanuit diens verwachtingen en geleerde vooroordelen over hoe het plaatje eruit zou moeten zien. Dit gebeurt op basis van diens ervaring en is bovendien tijd en situatieafhankelijk. Arnheim beargumenteert juist dat perceptie van realisme, na toepassing van de Gestaltregels, vooral wordt beïnvloed door de wijze waarop de uitbeeldende stijl (schilderkunst, tekening, foto en zo meer) representeert wat we weten van dat object (conceptuele realiteit) in vergelijking tot hoe het object er uitziet (perceptuele realiteit). Hij stelt net als Gombrich, dat dit intern bepaald wordt door de waarnemer zelf.

De laatste verklaring impliceert zelfs dat de waarnemer voordeel kan hebben van minder realiteit in een representatie omdat het informatiereductieproces tot de organisatie van kenmerken in vormen en patronen daarmee is ingekort.

Informatieverwerking en geheugenopslag volgens de Cognitieve Theorie van Multimedia Leren

Volgens enkele onderzoekers (Anglin, Vaez & Cunningham, 2004; Mayer & Moreno, 1998; Sweller, 2005) draagt kennis van geheugenprocessen bij aan het ontwerpen van een goede instructie ondermeer doordat ze de ontwerper bewust maken van de mogelijkheden en beperkingen van het geheugen, de (in)effectiviteit van visuele representaties in instructies alsook de mogelijkheden voor betekenisvol leren. De meest actuele theorie over geheugenprocessen is de Cognitive Theory of Multimedia Learning (CTML) (Mayer, 2005) die gebaseerd is

op Paivio's (1986) dual coding theory, Baddeley's (1992) model van het werkgeheugen, Sweller's (Chandler & Sweller, 1991; Sweller, Chandler, Tierney & Cooper, 1990) cognitive load theory, Wittrock's (1989) generative theory en Mayer's (1996) SOI (Select Organize Integrate) model voor betekenisvol leren.

Deze theorie, gericht op betekenisvol leren, gaat uit van een drietal principes. Het eerste principe is dat informatie de lerende bereikt via twee kanalen namelijk één dat visueel is en één dat auditief is. Het tweede principe is dat elk kanaal onderworpen is aan een beperkte verwerkingscapaciteit. Het derde principe is dat informatieverwerking een proces is van actief leren dat gepaard gaat met de uitvoer van een gecoördineerde set van cognitieve processen tijdens het leren. Deze processen zijn onder te verdelen in fasen horende bij een geheugentype: het selecteren van relevante informatie in het sensorisch geheugen, het organiseren van het geselecteerde in het werkgeheugen en vervolgens het integreren van deze nieuwe informatie met al aanwezige voorkennis in het lange termijngeheugen.

De CTML gaat uit van vijf cognitieve processen: het selecteren van relevante woorden uit de gepresenteerde tekst of verbale boodschap, het selecteren van relevante beelden uit gepresenteerde visuele representaties, het organiseren van geselecteerde woorden in een coherente verbale representatie, het organiseren van geselecteerde beelden in een coherente picturale representatie, en tot slot het integreren van verbale en picturale representaties evenals voorkennis. Het eindresultaat is een geïntegreerde representatie van de kennis bij de lerende.

Het principe van een beperkte verwerkingscapaciteit binnen elk kanaal dwingt tot het maken van keuzes in de relevantie van inkomende informatie: waar wel of geen aandacht aan wordt besteed, de mate waarin relevante relaties tussen het geselecteerde worden gelegd en de mate waarin dat vervolgens met bestaande kennis wordt geïntegreerd.

Het principe van actief leren vindt plaats als de lerende cognitieve processen toepast op inkomende informatie met als doel zinvolle informatie-eenheden te genereren. De uitkomst is de constructie van een coherente mentale representatie. Actief leren kan worden gezien als het proces waarin mentale modellen (kennisstructuren) worden samengesteld.

In relatie tot het hier gepresenteerde onderzoek kan het mentaal model worden geïnterpreteerd als kennis over een bepaald domein die een persoon heeft. Vanuit instructie kan deze kennis over een bepaald domein worden gezien als een veronderstelde eigenschap die de persoon al dan niet beheerst, het construct (Cronbach en Meehl, 1955) zoals eerder omschreven in dit hoofdstuk en in hoofdstuk 1. Vertaald naar het hier gepresenteerde onderzoek heeft de lerende zich een mentaal model gevormd bij het actief bestuderen van het kennisdomein anatomie van de rat dat equivalent behoort te zijn aan het gedefinieerde construct. Dit construct is het eindproduct van de instructie en de gestelde leerdoelstellingen en is aan toetsing onderhevig. Dit betekent dat de leerdoelstellingen een afspiegeling dienen te zijn van de definitie van het construct.

De CTML-theorie vertoont op een aantal onderdelen overeenkomsten met de perceptietheorieën van Gombrich (1969) en Arnheim (1974). Allereerst de actieve rol van de waarnemer/ de lerende in de verwerking en interpretatie van de geselecteerde informatie. Vervolgens het principe van informatiereductie, het selecteren van die informatie die relevant is voor verdere verwerking. Ook hierin is een rol voor realisme vastgelegd aangezien het selecteren van informatie om vervolgens met actief leren tot zinvolle informatie-eenheden te komen impliceert dat er een informatiereductieproces plaatsvindt hetgeen ook effecten heeft voor de mate van realiteit die vereist is voor het verkrijgen van (een) coherente mentale representatie(s).

Tegen de achtergrond van deze beschrijving van de cognitieve processen wordt in de volgende paragrafen dieper ingegaan op visualisaties en onderzoeken verricht met visualisaties.

Visuele representaties in relatie tot cognitieve processen

Indeling visuele representaties

Visuele representaties zijn onder te brengen in een aantal classificaties, te weten classificaties op basis van de vorm van de visuele representaties, de functie van de visuele representaties en het type kennis dat de lerende opdoet. Naast het vormen van een mentaal model kan de lerende ook een mentale animatie opnemen uit een visuele representatie. De volgende subparagrafen gaan dieper in op de verschillende classificaties.

(A) Classificaties naar vorm

Braden (1996) onderscheidt visuele representaties in vijf categorieën van onderzoek, te weten: (1) semiotiek en film/video conventies, (2) tekens, symbolen en iconen, (3) multi-image, (4) grafische representaties en (5) afbeeldingen en illustraties. De visuele representaties zoals bedoeld in het hier gepresenteerde onderzoek behoren tot de laatste categorie.

Knowlton (1966) differentieert vervolgens de categorie afbeeldingen en illustraties in drie groepen, te weten realistische, analoge en logische representaties. Deze differentiatie is vergelijkbaar met Alesandrini's classificatie (Alesandrini, 1984) van representatieve, analoge en arbitraire representaties.

Realistische of representatieve illustraties vertonen een fysieke gelijkenis met het referentieobject of concept. Voorbeelden daarvan zijn foto's van objecten die de realiteit benaderen maar ook schema's van dezelfde objecten die door het weergeven van contouren een duidelijke fysieke gelijkenis met het referentieobject vertonen. In figuur 2.1 worden, vanuit het kennisdomein van de rat dat van toepassing is in het hier gepresenteerde onderzoek, een aantal voorbeelden van de rattenmaag weergegeven die een fysieke gelijkenis met het referentieobject vertonen.

Figuur 2.1 Overzicht van representatieve weergaven van de rattenmaag, al dan niet in relatie tot omliggende objecten

In alle vier illustraties is de maag duidelijk zichtbaar en herkenbaar met vormen die fysieke gelijkenis(en) vertonen met het referentie object. Toch verschillen de illustraties van elkaar in hun mate van realiteitsweergave wat zich vooral kenmerkt in de mate waarin details worden vertoond. Globaal genomen zijn hierin twee aspecten zichtbaar. Allereerst de mate van detailweergave door

representaties in fotovorm en representaties in schemavorm weer te geven. De schematische weergave bevat veel minder visuele kenmerken dan de fotoweergave terwijl beide vormen voorzien in voldoende relevante informatie over het object ten behoeve van de herkenbaarheid ervan door kenmerken als vorm en positie in relatie tot andere objecten. Ten tweede de mate van detailweergave waarbij het object in een illustratie wordt weergegeven in de context waarin deze ligt binnen het reële object zoals door de student geobserveerd en een illustratie waarbij het object wordt weergegeven zonder de context van het reële object. De maag als afzonderlijk object weergegeven biedt informatie over dat object maar geen informatie over de wijze waarop de maag is gepositioneerd in de rat (het studieobject) en diens relatie tot andere daaraan verbonden onderdelen zoals weergegeven met context.

Deze vier weergaven zijn indicierend voor de mogelijke varianten in de weergaven van de werkelijkheid en daarmee ook illustratief voor de eerder in dit hoofdstuk veronderstelde rangschikking in weergaven van meer naar minder realistisch. Figuren a en c zijn realistischere weergaven van een object in vergelijking tot figuren b en d die hetzelfde object weergeven, element realisme. Figuren a en b bevatten meer informatie in vergelijking tot c en d omdat ze naast het object ook de andere objecten in de omgeving weergeven, de contextuele informatie waardoor ook informatie beschikbaar is over de grootte van het object in relatie tot de andere objecten, element context.

Analoge illustraties vertonen iets anders dan het referentieobject maar zijn overeenkomstig doordat ze een gelijkenis suggereren. Een voorbeeld daarvan is het uitbeelden van anatomische objecten zoals de maag door symbolen zoals een ovaal. Er is dan toegevoegde tekst noodzakelijk zodat de waarnemer door de combinatie tekst en illustratie correct waarneemt en het geheel begrijpt. In de volgende weergave van de rattenmaag (figuur 2.2) is een verklarende beschrijving van of verwijzing noodzakelijk.

Figuur 2.2 Een analoge illustratie van een maag als een ovaal waarbij de stroomrichting is aangegeven

Met een beschrijving over de maag en de illustratie wordt duidelijk wat deze is, de oriëntatie ervan en de locatie van de in- en uitgang. Verbale informatie is bij bovenstaande illustratie noodzakelijk om het object te herkennen.

Logische of arbitraire illustraties vertonen op geen enkele wijze gelijkenis met het referentieobject maar geven een organisatorische samenstelling of lay-out weer die een bepaalde conceptuele of logische relatie van de onderdelen in een illustratie benadrukken. Een voorbeeld daarvan is een weergave van het niersysteem van de rat dat in figuur 2.3 is weergegeven.

Figuur 2.3 Van ventraal aanzicht: een logische of arbitraire illustratie waarin de verbindingen van de aorta (A) met de nieren (N) en de bijnieren (B) zijn getekend, evenals de verbindingen van de ovaria (O) met de ongepaarde vagina (V).

Een andere classificatie naar vorm is afkomstig van Anglin, Vaez en Cunningham (2004), die een onderscheid maken tussen statische representaties, zoals tekeningen, en geanimeerde representaties, zoals films. In het hier gepresenteerde onderzoek is het reële object in statische vorm het uitgangspunt met als gevolg dat alle weergaven zijn beperkt tot statische representaties.

In recenter onderzoek naar multimediaal leren en de rol van visuele representaties (Ainsworth, 2006; Hegarty, 2005; Schnotz, 2005) wordt het

onderscheid gemaakt tussen externe en interne visuele representaties dat in hoofdstuk 1 genoemd is. Een externe representatie refereert naar de visuele representatie van het object zelf. Een interne representatie refereert naar het mentaal model die de waarnemer heeft van de visuele representatie van het referentieobject of het object zelf. Seels (1994) geeft aan dat iemand die visueel leert ook denkt in termen van beelden, meer specifiek mentale beelden. In relatie tot de instructie in het hier gepresenteerde onderzoek is aannemelijk dat deze mentale beelden bij elkaar een substantiële bijdrage leveren aan de vorming van de visuele component van het construct (kennis over een bepaald domein) zoals eerder in paragraaf 2.1 en hoofdstuk 1 is uitgelegd.

(B) Classificaties naar functie

Duchastel en Waller (1979) kennen illustraties een zevental functies toe. Ze onderscheiden een descriptieve functie voor een fysieke weergave van het referentieobject; een expressieve functie met als doel indruk te maken op de lezer; een constructieve functie om te laten zien hoe de verschillende delen van een systeem één geheel vormen; een functionele functie met als doel het proces of de organisatie van een systeem te laten zien; een logisch-mathematische functie voor het tonen van een mathematisch concept; een algoritmische functie om mogelijke acties zichtbaar te maken en tot slot een data-display functie voor een visuele vergelijking en toegankelijkheid van data zoals grafieken. In relatie tot het hier gepresenteerde onderzoek waarbij de lerende de rat bestudeert in diverse dissectielagen met daarbij behorende afbeeldingen, dienen de afbeeldingen functioneel te zijn aan het, door de lerende, te vormen construct. Een afbeelding vervult een functie bij het vormen van een correct construct en in het hier gepresenteerde onderzoek naar het leren en het toetsen van dat construct zijn een drietal functies van belang. Een descriptieve functie waarbij de fysieke weergave van het reële object is weergegeven zodat de lerende anatomische structuren kan herkennen aan voornamelijk vorm en kleur (het eerste element van het construct), en deze ook benoemen (het tweede element van het construct). Door een constructieve en/of een functionele functie toe te kennen aan een afbeelding wordt voldaan aan het derde element van het te leren construct, het kunnen aangeven van de ruimtelijke relaties tussen deze structuren rekening houdend met de relatieve grootte van elke structuur in relatie tot andere structuren waarmee samen een geheel wordt gevormd. Afbeeldingen a en b in figuur 2.1 zijn daar voorbeelden van. Hiermee wordt een relatie gelegd tussen het te leren construct en de vorm en functie van de afbeelding. De twee functies constructief

en functioneel, zouden in principe ook kunnen worden behaald met afbeeldingen zoals weergegeven in de figuren 2.2 en 2.3.

Levie en Lentz (1982) stelden voor een classificatie te maken naar de mate waarin een illustratie effect heeft op een lerende. Dit leverde vier functies op. Een aandachttrekkende functie met als doel dat de waarnemer zich specifiek richt op de afbeelding; een affectieve functie waarbij een gevoel/emotie wordt opwekt of versterkt; een cognitieve functie met als doel het leerproces te bevorderen door bevordering van het begrip en/of de retentie of het voorzien in additionele informatie; de compenserende functie bedoelt ter ondersteuning van zwakke lezers of lerenden. De classificatie van Levie en Lentz (1982) is voor het hier gepresenteerde onderzoek toepasbaar op het reële object als studieobject. Er is sprake van een affectieve functie doordat gevoel/emotie wordt opgewekt doordat de waarnemer wordt geconfronteerd met een dier dat kort daarvoor levend was, aaibaar is en een geur heeft. De instructiesituatie voorziet in een aandachttrekkende en een cognitieve functie in het leerproces. De student wordt verplicht een afbeelding te produceren waardoor deze de aandacht moet richten op relevante onderdelen/systemen van het reële object hetgeen vervolgens ertoe leidt dat het begrip over het reële object wordt bevorderd. Ook binnen deze classificatie is de relatie gelegd tussen te leren construct, de functie van het object alsook de functie en vorm van de afbeelding. In de onderwijspraktijk die voor het onderzoek de beginsituatie vormt, is als instructie- en tevens toetsinstrument gekozen voor de productie van representatieve afbeeldingen door studenten zoals weergegeven in figuur 2.1b.

(C) Classificatie naar typen kennis

Naast de classificatie van visuele representaties naar vorm en functie is er ook een classificatie naar type kennis. Hegarty (2005) onderscheidt verschillende typen kennis die een lerende op kan doen aan de hand van afbeeldingen waarbij zij zich baseert op de CTML-theorie. Meer specifiek benoemt zij hierin diagrammen. Diagrammen zijn hierin gedefinieerd als speciale afbeeldingen die fysieke gelijkenissen vertonen met het referentieobject waarbij naast de fysieke relaties ook de fysieke bewegingsrichting van of in het systeem wordt weergegeven. Figuur 2.4 geeft twee diagrammen die beantwoorden aan de definitie van Hegarty.

Figuur 2.4 A. Diagram van een band en treksysteem met annotatie van de richting. B. Diagram van een treksysteem. Uit: *The Cambridge handbook of Multimedia learning*. Richard Mayer 2005. Chapter 28. *Multimedia learning about physical systems*. Mary Hegarty

In de diagrammen onderscheidt Hegarty ten eerste kennis over de statische configuratie van het systeem zoals de delen van het systeem, uit welk materiaal ze bestaan en hoe deze met elkaar verbonden zijn en ruimtelijk zijn geconfigureerd. Figuren 2.1 A en 2.1 B zijn daar voorbeelden van. Ten tweede onderscheidt zij kennis over de wijze waarop de componenten bewegen en elkaar beïnvloeden, het causale gedrag van het systeem. Dit betekent onder andere kennis van de dynamische krachten en hoe die elkaar in het systeem beïnvloeden. Figuur 2.4 A is daar een voorbeeld van. Als derde type is er kennis over de functies van het systeem, meer specifiek waar het systeem toe dient en hoe de structuur en het gedrag ertoe bijdragen dat deze functies worden vervuld. Deze drie kennistypen zou de student ook opdoen in de aangegeven volgorde. Hegarty legt hiermee de relatie tussen typen kennis die worden opgedaan en de daartoe vereiste vorm en functie. De kennistypen kunnen ook worden vertaald naar het te leren construct in het hier gepresenteerde onderzoek waarbij vooral de statische configuratie van het systeem van belang is in de representatie.

In het in hoofdstuk 1 genoemde PSE-programma beschrijft Dwyer (1978) in de eerste fase van zijn onderzoekstraject naar visueel leren het bestaan van verschillende typen kennis die hij relateert aan leertaken en vervolgens aan toetstaken. Hij onderscheidt vier typen kennis die hiërarchisch zijn geordend van

simpel naar complex. In toenemende moeilijkheidsgraad gaat het om: (1) kennen van terminologie en feiten, (2) identificeren van locaties en posities, (3) constructie en/of begrip van relaties en (4) probleemoplossende activiteiten. Dwyer baseert zich op de theorie van Gagné (1965) waarin leertaken voor het leren van intelligente vaardigheden hiërarchisch geordend zijn. Op basis van Gagné's theorie over leercondities ontwierp Dwyer (1978) zowel de instructie als de toets van het PSE-programma. In paragraaf 4 van dit hoofdstuk wordt daarop verder ingegaan.

Het gedefinieerde construct, equivalent aan de leerdoelstellingen, in het hier gepresenteerde onderzoek komt overeen met het eerste kennistype van Hegarty, de statische configuratie van het systeem en met de eerste twee typen kennis zoals gesteld door Dwyer. In paragraaf 3 over instructies wordt de relatie leertaken en visualisaties nader uitgewerkt. In hoofdstuk 4 zal na de evaluatie van de bestaande instructie en vooraf aan de ontwikkeling van de toets blijken dat de docenten de gestelde leerdoelstellingen en dus ook de definitie van het te leren construct willen uitbreiden met functionele kennis van het systeem waarbij constructie en/of begrippen van relaties relevant zijn.

(D) Mentaal model en mentale animatie

Aan de hand van de typen kennis die Hegarty (2005) classificeert, gaat zij vervolgens een stap verder in de typeringsmogelijkheden van visualisaties. Zij benoemt een interne representatie van een fysiek systeem als een mentaal model. Daarbij gaat zij ervan uit dat deze wordt gebruikt ter aanduiding van een representatie die gelijkwaardig is aan zijn referentieobject. Vervolgens benoemt zij een mentale animatie. Een mentale animatie is het proces waarbij wordt afgeleid van een statische visuele representatie hoe een object beweegt of een fysieke stroming verloopt. Zij relateert hieraan de kennis die de lerende opdoet over de causale relaties of causale gebeurtenissen in het gedrag van het fysiek systeem. Lerenden visualiseren bij het bestuderen van een statisch mechanisch diagram, zoals het spoelsysteem van een toilet, in gedachten opeenvolgende causale acties die vereist zijn voor de werking van het weergegeven mechanisch systeem. Het vermogen tot dergelijke mentale animaties is echter beperkt en lerenden met weinig domeinkennis hebben erg veel moeite met dit visualisatieproces en zijn daarin ook weinig accuraat (Hegarty & Steinhoff, 1997). Toevoeging van coderingen als pijlen die de bewegingen aangeven stimuleren lerenden het verband te leggen met de diverse functies van het realistische object en zijn mechanismen (Heiser & Tversky, 2002). Ook in deze is de relatie vorm, functie en te leren kennistype herkenbaar.

Vertaald naar het hier beschreven onderzoek over de anatomie van de rat is het spijsverteringssysteem een voorbeeld. Een statische visuele representatie van een spijsverteringssysteem, in de vorm van een diagram, kan bij een lerende niet alleen een mentaal model creëren maar vervolgens ook een mentale animatie omdat de lerende aan de hand van dit diagram kennis opdoet over de processen die elkaar in het systeem beïnvloeden en hun dynamiek. Daarmee wordt causale kennis opgedaan over het spijsverteringssysteem. In hoofdstuk 4 bij de herdefiniëring van het te leren construct wordt dit aspect opgepakt.

Bovenstaande categorieën laten ten aanzien van visuele representaties een grote diversiteit zien in de functies die ze kunnen vervullen, in de typen kennis die de lerende daarmee kan opdoen en in de mentale modellen die daarmee kunnen worden gecreëerd.

Voor het hier gepresenteerde onderzoek biedt deze inventarisatie nog onvoldoende informatie om keuzes te maken in representaties die geschikt zijn voor de instructie over het bedoelde construct of voor de toetsing van dat construct. In de volgende paragrafen worden getracht hier meer duidelijkheid in te krijgen.

Effectiviteit van illustraties in teksten

Levie en Lentz (1982) bestudeerden 55 experimenten waarin teksten zonder illustraties werden vergeleken met teksten met illustraties. Hun centrale onderzoeksvraag was of illustraties positief bijdragen aan het leren van teksten. Zij destilleerden daar negen algemene conclusies uit waarvan twee relevant zijn voor het hier beschreven onderzoek. De eerste is dat voor het effectief gebruik maken van complexe illustraties expliciete instructie vereist is. De tweede stelt dat het aanbieden van illustraties over het algemeen nuttiger is dan dat de lerende zelf een eigen beeldvorming ontwikkelt naar aanleiding van de tekst of van een op eigen initiatief geproduceerde tekening.

Hun (Levie & Lentz, 1982) conclusie dat lerenden expliciet instructie moeten krijgen voor het effectief gebruik maken van complexe illustraties, is gebaseerd op de stelling dat lerenden illustraties in principe oppervlakkig bekijken zonder de verwachting daar bruikbare informatie in te vinden. De details in de illustraties worden regelmatig over het hoofd gezien en er is sturing voor nodig om de lerende daarop te attenderen. Zowel Gombrich (1969) als Arnheim (1974) geven aan dat tijdens de perceptie er een selectie plaatsvindt van informatie waarbij de

waarnemer een actieve rol heeft in het bepalen welke informatie relevant is en het begrijpen van deze informatie. De CTML theorie wijst op de beperkte verwerkingscapaciteit van het visuele en auditieve kanaal wat de lerende dwingt te kiezen waaraan wel of geen aandacht wordt besteed, welke relaties tussen geselecteerde informatie als relevant worden beschouwd en in welke mate dit wordt geïntegreerd met bestaande kennis. Deze effecten van de beperkte verwerkingscapaciteit ondersteunen daarmee de conclusie dat bij het aanbieden van een complexe illustratie, sturing van lerenden bijdraagt aan het selecteren van de juiste informatie. Sturing van de lerende waarborgt dat de lerende de correcte informatie tot zich neemt en daarmee het juiste mentale model creëert. Door in een instructie sturing te geven wordt de aandacht van elke lerende op eenduidige wijze gericht op relevante informatie wat ertoe kan bijdragen dat de lerenden collectief de juiste informatie tot zich nemen voor het vormen van een eigen mentaal model.

De tweede conclusie dat het aanbieden van illustraties over het algemeen nuttiger is dan een eigen beeldvorming, moet gezien worden in de context van de door de auteurs geanalyseerde onderzoeken waarin de leereffecten van teksten zonder illustraties en teksten met illustraties werden vergeleken aan de hand van de opdracht aan de lerenden om een eigen mentaal beeld te vormen van het beschrevene of zelf tekeningen te produceren. De uitkomsten van deze onderzoeken waren niet consistent. Een aantal onderzoeken vertoonde resultaten waarin lerenden wel in staat bleken een correct mentaal beeld te vormen maar een aantal onderzoeken vond dat ook niet. Voornamelijk jonge kinderen bezitten niet de vaardigheden om hulpmiddelen, zoals het zelf tekenen ter uitbeelding van de tekst, in te zetten voor een beter begrip daarvan. Bovendien speculeren Levie en Lentz (1982) dat een hulpmiddel als het tekenen mogelijk bijdraagt aan de retentie van de informatie maar minder aan het begrijpen van het bedoelde kennisdomein. Zij stellen kortweg dat het aan lerenden aanbieden van relevante illustraties behorend bij de teksten in alle gevallen betrouwbaarder is. Voor het hier gepresenteerde onderzoek is de functie van het produceren van een tekening het weergeven van een observatie van een concreet leerobject. Bij Levie en Lentz is de functie van het tekenen het weergeven van de inhoud van de tekst zoals de lerende zich deze voorstelt tijdens het lezen daarvan, een veel abstractere aangelegenheid. Overigens betreffen veel van hun conclusies jonge kinderen waardoor deze conclusie met de nodige voorzichtigheid moet worden gedaan. De conclusie van Levie en Lentz is hierdoor niet logischerwijs één op één over te nemen maar laat wel zien dat het nut van tekenen niet altijd vanzelfsprekend is.

Effectiviteit van teksten bij diagrammen

Waar in verschillende onderzoeken vooral wordt uitgegaan van het leren van teksten en de bijdrage van illustraties aan het leerproces als deze worden toegevoegd, past Hegarty de omgekeerde benadering toe. Daarin wordt de vraag gesteld of een diagram van een fysiek systeem voldoende is om zich in het bijzonder causale kennis over dat systeem eigen te maken of dat toevoeging van tekst daaraan substantieel bijdraagt. De conclusie die Hegarty (2005) trekt is dat aan diagrammen toegevoegde verbale instructie effectief kan bijdragen aan het leerproces. Door de optelsom van informatie, zowel verbaal als visueel, wordt informatie dubbel aangeleverd en dus beter onthouden. Vervolgens kan verbale instructie de lerende sturen in het verwerken van de relevante informatie in het diagram. Verbale instructie kan ook de stappen in het systeem, die weergegeven zijn in het diagram, voor de lerende ordenen. Tot slot kan verbale instructie informatie verschaffen die niet is weergegeven in het diagram.

Kortom, een tekstuele beschrijving toegevoegd aan een diagram kan de aandacht van de lerende sturen, richting geven aan een mentale animatie alsook informatie bieden over niet te visualiseren aspecten.

Samengevat kan over cognitieve processen en de rol van betrokken visuele representaties het volgende worden gesteld.

Vanaf het percipiëren van een object tot de cognitieve verwerking naar een mentaal model van dat object is de waarnemer continue bezig de juiste informatie te selecteren en daar zinvolle informatie-eenheden van te maken. Als zodanig wordt getracht effectief met informatie om te gaan wat vooral noodzakelijk is vanwege de beperkte verwerkingscapaciteit van het geheugen. De waarnemer is actief betrokken bij dit proces zowel tijdens perceptie als tijdens de cognitieve verwerking voor het vormen van een mentaal model en heeft als zodanig invloed op dit proces. Dit actief leren wordt beschouwd als het proces waarin mentale modellen (kennisstructuren) worden samengesteld en in het verlengde daarvan als het eindresultaat van het leerproces ofwel het eindproduct van de instructie. De ontwikkelde mentale modellen worden verondersteld de kennis te representeren die de persoon heeft opgedaan over het kennisdomein. Deze kennis over een bepaald domein is de veronderstelde eigenschap, het construct die de te toetsen persoon al dan niet beheerst. Met dit gegeven wordt de vraag gesteld hoe vanuit instructie kan worden beïnvloed dat lerenden tot beoogde constructen komen en hoe dit eindresultaat kan

worden getoetst. Verondersteld wordt dat de leerdoelstellingen van een instructie equivalent moeten zijn aan de constructdefinitie. Vervolgens geeft de literatuur hierin het belang aan van het sturen van het leerproces, ondermeer door expliciete instructies, zoals expliciete aanwijzingen en voorgeschreven taken of handelingen, om de aandacht van de lerenden te richten op relevante informatie waardoor de juiste selectie kan plaatsvinden. In de volgende paragraaf over instructie wordt hierop dieper ingegaan.

Vervolgens zijn visuele representaties besproken door classificaties naar vorm, naar functie en naar type kennis dat kan worden opgedaan. Daarbij is ook de relatie tussen te leren construct, vorm, functie en type kennis gelegd die in de volgende paragrafen nader wordt beschouwd. Ten aanzien van de vorm zijn de representaties beperkt tot de categorie afbeeldingen en illustraties die vervolgens differentieerbaar zijn naar varianten in de weergave van de werkelijkheid. Voor het hier gepresenteerde onderzoek zijn daar geen keuzes in gemaakt. Wel wordt beperkt tot statische representaties, intern dan wel extern. Voorts is er, gericht op het hier gepresenteerde onderzoek, een speciale groep representaties vermeld, diagrammen genaamd, die de fysieke processen in een systeem weergegeven. Van deze groep diagrammen is aangegeven dat de lerende zich een mentale animatie kan vormen door, vanuit de statische illustratie, de opeenvolgende processen die plaatsvinden in het systeem te visualiseren. De typen kennis die een lerende opdoet bij het bestuderen van ondermeer diagrammen zijn benoemd als kennis over de statische configuratie van het systeem, kennis over het causale gedrag ervan alsook kennis over de functies ervan. Deze typen kennis in relatie tot instructie en toetsing worden in de volgende paragrafen verkend. De inventarisatie naar vorm van representaties, de te vervullen functie en de kennis die daarbij moet worden opgedaan voor het hier gepresenteerde onderzoek, kan in een voorlopig overzicht worden geplaatst zoals weergegeven in tabel 2.1.

Tabel 2.1 *Rangschikking representaties ten behoeve van gepresenteerd onderzoek*

Categorie	Type	Beschrijving	Mogelijke te vervullen functies	Mogelijk te verkrijgen kennistype	
Realistisch	Foto binnen context (figuur 2.1a)	Vertoont in fotovorm het geheel waarbinnen een (deel)object zich bevindt	Descriptief, constructief, functioneel	▪ Statische configuratie ▪ Causaal gedrag systeem ▪ Functies systeem	
	Schema binnen context (figuur 2.1b)	Vertoont in schematische vorm het geheel waarbinnen een (deel)object zich bevindt	Descriptief, constructief, functioneel	▪ Statische configuratie ▪ Causaal gedrag systeem ▪ Functies systeem	
	Foto (figuur 2.1c)	Vertoont in fotovorm het (deel)object op zichzelf los van het geheel	Descriptief	Statische configuratie	
	Schema (figuur 2.1d)	Vertoont in schematische vorm het (deel)object op zichzelf los van het geheel	Descriptief	Statische configuratie	
Diagram	Diagram (figuur 2.4)	Vertoont fysieke gelijkenis met het geheel en de onderlinge fysieke relaties tussen deelobjecten	Beperkt descriptief, constructief, functioneel	▪ Statische configuratie ▪ Causaal gedrag systeem ▪ Functies systeem	
Analoog	Analoog (figuur 2.2)	Suggereert een gelijkenis met het object waarbij toegevoegde tekst een vereiste is	Beperkt descriptief, Functioneel	Beperkt statistische configuratie	Maag, ventraal aanzicht
Logisch	Logisch (vergelijkbaar met figuur 2.3)	Vertoont geen gelijkenis met het reële object maar wel een organisatorische samenstelling of lay-out van gerelateerde (deel)objecten	Constructief, functioneel	Causaal gedrag systeem	

In de volgende paragrafen wordt deze classificatie verkend in relatie tot het construct, de instructie en toetsen.

2.3 INSTRUCTIES MET REPRESENTATIES (VISUEEL LEREN)

Met de informatie uit de vorige paragraaf wordt in deze paragraaf de literatuur toegespitst op de voor dit onderzoek geldende instructie voor het behalen van de leerdoelstellingen in relatie tot visuele representaties.

Sturing geven aan het leerproces met taakgerichte opdrachten

Bij het percipiëren van een object of diens referent, een plaatje, vindt er een eerste selectie van informatie plaats, waaronder de visuele kenmerken van het object. Vertrekkend vanuit het principe dat de waarnemer actief betrokken is bij de selectie van informatie uit een plaatje en de interpretatie ervan, is er onderzoek (Levie, 1987; Stern & Robinson, 1994; Bultler & Mc Kelvie, 1985) verricht naar de wijze waarop een waarnemer aandacht besteedt aan een plaatje of een bepaald onderdeel in dat plaatje. Veel van dit onderzoek is gebaseerd op het principe van selectieve aandacht waarbij de onderzoeker bijhoudt waar een waarnemer naar kijkt in een plaatje op elk moment tijdens het scannen ervan. Factoren die van invloed zijn op het bepalen van de belangrijke gebieden van een plaatje tijdens het scannen zijn (Levie, 1987) (a) de verwachtingen van de waarnemer, (b) het doel van de waarnemer (exploratief of taakgericht) en (c) de kenmerken van de waarnemer zoals leeftijd, oogsterkte, cultuur en expertise in het decoderen van het plaatje. Butler en McKelvie (1985) ondersteunen deze stelling naar aanleiding van onderzoek dat zich richt op het doel van de waarnemer als deze een plaatje scant. Zij hebben de herkenning van plaatjes onderzocht als vooraf aan de kandidaten was verteld op welke vormen ze moesten letten. De kandidaten kregen eerst de opdracht om plaatjes te bekijken op basis van bepaalde kenmerken. Daarna werd de herkenning getest op kenmerken waar de aandacht expliciet op was gericht, kenmerken waar de aandacht niet op gericht was en nieuwe plaatjes die aan de test waren toegevoegd. Het resultaat was dat plaatjes met geëxpliciteerde kenmerken vaker werden herkend. De onderzoekers concludeerden dat voor het leren van plaatjes (a) doelbewuste aandacht vereist is en (b) dat het geven van taakgerichte opdrachten bijdraagt aan het herkennen. Deze conclusie is voor het hier gepresenteerde onderzoek belangrijk omdat een taakgerichte opdracht een positieve bijdrage kan leveren door de aandacht

specifiek te sturen en als zodanig te richten op die visuele aspecten die van belang zijn voor de kennisontwikkeling in de instructie. Daar kan vervolgens op worden ingespeeld. Dat dit noodzakelijk is bevestigt Pettersson (1993) die concludeert dat bij vergelijking van teksten zonder illustratie en teksten met illustratie, visuele representaties onderhevig zijn aan individuele interpretatie maar dat toch in ruime mate lezers datgene zien wat hen wordt verteld te zien. Ook hierin wordt het belang van sturing aan het leerproces aangestipt.

Binnen het biologieonderwijs en zo ook in de bestaande instructie van het hier beschreven onderzoek, wordt het maken van een representatieve tekening van het reële object op basis van observaties als een geschikte methodiek beschouwd om studenten een goede mentale representatie te geven van het object. De productie van de tekening kan als een taakgerichte opdracht worden gezien. In meerdere onderzoeken naar de toegevoegde waarde van tekenen als instructiemethode bleek deze geschikt voor sturing in de selectie van relevante visuele kenmerken en/of relaties (Van Meter, Aleksic, Schwartz & Garner, 2006; Van Meter & Garner, 2005; Van Meter, 2001; Lansing, 1984, 1981). In studies waarin de lerenden werd verteld hoe ze aangeboden visuele informatie moesten verwerken, nam de prestatie toe (Kosslyn, 1980). Dat deze overtuiging klopt wordt bevestigd in onderzoek naar het verkrijgen van een concreet mentaal beeld van een object door het maken van een tekening (Van Meter & Garner, 2005; Lansing, 1984, 1981). Volgens Lansing (1984) draagt het tekenen van een object ertoe bij dat de tekenaar/waarnemer een goede mentale representatie van het te bestuderen object ontwikkelt en de visuele kenmerken van het object leert kennen. Daar zijn wel voorwaarden aan verbonden, te weten (a) de waarnemer moet voor het maken van de tekening worden geïnstrueerd over de belangrijke visuele kenmerken en de belangrijke relaties tussen de kenmerken en (b) de waarnemer moet het object meerdere malen tekenen. Dit onderzoek is verricht met 145 kinderen uit het basisonderwijs, die de opdracht kregen om een tweedimensionaal plaatje, zijnde een abstract geschematiseerd plaatje dat niet naar een specifiek object refereert, te bestuderen en daarna te tekenen. De resultaten van het onderzoek bevestigden Lansing's verwachting dat tekenen als instructiemiddel bijdraagt aan het verwerven van een correct mentaal model mits het verwervingsproces wordt gestuurd door instructie. Hoewel dit onderzoek is uitgevoerd met kinderen is Lansing ervan overtuigd dat soortgelijke principes ook zouden gelden voor volwassenen.

Van Meter (2001) heeft onderzoek verricht naar het maken van een representatieve tekening als leerstrategie voor het leren van een tekst met vijfde- en zesdejaars studenten. De studenten kregen allemaal een twee pagina's tellende tekst over het centrale zenuwstelsel. Daarbij waren er vier condities waarvan drie experimentele condities waarin studenten na een korte tekentraining verschillend ondersteund werden tijdens het maken van een tekening: (a) een experimentele conditie waarin deelnemers tekeningen produceerden na elke pagina, (b) een experimentele conditie waarin deelnemers tekeningen produceerden na elke pagina en deze vervolgens vergeleken met de illustraties die ze daarna ontvingen, (c) een experimentele conditie waarin deelnemers tekeningen produceerden na elke pagina, deze vergeleken met de aangeboden illustraties en vervolgens mondeling vragen beantwoorden die richtinggevend waren in het vergelijken van de geproduceerde tekeningen en aangeboden illustraties en (d) een controleconditie waarin niet werd getekend maar de studenten illustraties kregen. Het onderzoeksresultaat was dat het construeren van een tekening effectiever was dan het krijgen van een tekening. Nog belangrijker was dat in de condities waarin werd getekend de groep met de meeste ondersteuning, ofwel sturing tijdens het vergelijken van de tekening en de illustraties, het hoogst scoorde. Net als bij Lansing (1981) bleek additionele ondersteuning ten aanzien van belangrijke kenmerken en relaties het effectiefst.

Deze conclusies plaatsen de eerder beschreven algemene conclusie van Levie en Lentz (1982) (blz 33) dat het aanbieden van illustraties over het algemeen nuttiger is dan dat de lerende zelf zijn beeldvorming ontwikkelt, onder andere door te tekenen op eigen initiatief, in een ander licht. In de 55 onderzoeken die zij analyseerden kregen de participanten geen sturing of instructie in wat getekend moest worden. Dat heeft hoogstwaarschijnlijk de door hen verkregen resultaten beïnvloed en als zodanig ook hun uiteindelijke conclusie. Voorts onderstrepen deze conclusies dat de functionaliteit van het tekenen als instrumentarium in het leerproces belangrijk is want deze kan, mits goed ingezet, richting geven aan het leerproces. Dit betekent dat bij de inzet van een tekenopdracht duidelijk gespecificeerd dient te zijn *wat* getekend moet worden. Ervan uitgaand dat sturing van het leerproces in de selectie van relevantie informatie resulteert in een correct mentaal beeld, ondersteunt de hier gepresenteerde analyse de assumptie van Jones, Reed en Weyers (1994) in zoverre dat het maken van een tekening een geschikte onderwijsmethode is voor het ontwikkelen van een mentaal beeld. Deze dient wel als taakgerichte opdracht te worden gegeven waarbij wordt gestuurd

op relevante te tekenen visuele kenmerken. De traditie om te tekenen is dus voornamelijk verantwoord door het taakgerichte karakter van de tekenopdrachten waarbij de besproken onderzoeken (Van Meter, 2001; Lansing, 1984) aanduiden dat 'goede' taakgerichte opdrachten met deze tekencomponent het effectiefst zijn als er sprake is van enige training vooraf. Van Meter (2001) voegt daaraan toe dat de effectiviteit toeneemt door het toevoegen van andere componenten zoals het vergelijken van de geproduceerde tekeningen met vergelijkbare illustraties en/of het mondeling stellen van vragen over het onderwerp. Met deze informatie kan tekenen worden beschouwd als een (onderdeel van een) taakgerichte opdracht. Zo ook in het hier gepresenteerde onderzoek waarbij tijdens de instructie studenten niet alleen een tekenopdracht krijgen maar de objectonderdelen moeten benoemen en worden geïnstrueerd in de wijze waarop ze het object dienen te observeren en de belangrijke onderdelen te tekenen (zie blz. 1 en 19).

Naast het tekenen van een object zijn er ook andere taakgerichte opdrachten die bijdragen aan het construeren van een correcte mentale representatie van een object door sturing, zoals het maken van een sculptuur van het reële object (Waters, Van Meter, Perrotti, Drogo & Cyr, 2005), deelname aan een interactief dissectiesimulatieprogramma (Kinzie, Strauss & Foss, 1993) of het benoemen van objecten en/of objectkenmerken in illustraties (Loftus & Bell, 1975). Loftus en Bell lieten in een experiment studenten tekeningen zien. Eén groep kreeg van tevoren de opdracht om de details in de tekeningen te benoemen terwijl de andere groep geen opdracht kreeg. Het resultaat was dat de groep die de opdracht kreeg de details te benoemen, significant hoger scoorde op herkenning dan de groep die deze opdracht niet had gekregen. De verklaring van de onderzoekers voor dit verschil is dat herkenning voornamelijk gebaseerd is op het vermogen om te discrimineren tussen plaatjes wat begint met te bepalen welk gebied in een plaatje het meest onderscheidend is in vergelijking tot andere plaatjes. Het benoemen van details levert aan het leerproces een positieve bijdrage door daar sturing aan te geven.

Deze alternatieven bieden in het hier gepresenteerde onderzoek de mogelijkheid te onderzoeken of de instructie ook met andere taakopdrachten kan worden vormgegeven dan met tekenopdrachten. Voor het hier beschreven onderzoek is dit van belang omdat, zoals in hoofdstuk 1 is aangegeven, gezocht wordt naar een toetsvorm waarin niet langer gebruik gemaakt wordt van het beoordelen van geproduceerde tekeningen. De bestaande op het maken van tekeningen

gebaseerde instructie zal daardoor mogelijkerwijs niet meer aansluiten op de te ontwikkelen toetsmethodiek. Voor een andere invulling van de instructie met taakopdrachten uit het bovenbeschreven onderzoek worden de taakcomponenten van het kunnen benoemen van objecten en/of objectkenmerken in illustraties beschouwd als te onderzoeken alternatieven omdat deze passen binnen de onderwijssituatie waarin het reële object een belangrijk onderdeel vormt en in de toepassing een sturende werking heeft. Andere alternatieven zoals deelname aan een dissectiesimulatieprogramma komen minder in aanmerking omdat dissectie en observatie van het reële object blijvende onderdelen zijn van de instructie.

Samenvattend is het noodzakelijk om in instructie sturing te geven aan het leerproces voor het ondersteunen van de selectie van relevante informatie. Taakgerichte opdrachten zijn hierin een belangrijk instrument. Deze opdrachten zijn richtinggevend, maken effectieve selectie mogelijk, dragen bij aan het optimaliseren van de beperkte verwerkingscapaciteit tijdens het leerproces en dragen ook bij aan het verkrijgen van correcte mentale beelden van objecten die samen en in samenhang het te leren construct vormen.

In hoofdstuk 3 wordt ingegaan op geschikte instructiemethoden voor het in dit onderzoek te leren construct door vergelijking van twee instructies waar in de onderwijsmethodiek verschillende taakopdrachten zijn verwerkt waarvan één met tekenopdrachten en één met benoemingsopdrachten in aangeboden tekeningen.

Vorm en functie van visuele representaties in relatie tot instructie en toetsing

In de literatuur is het realiteitsgehalte van een representatie als belangrijk visueel kenmerk meerdere malen benoemd. Zoals te zien is in tabel 2.1 zijn er vele variaties in de mate waarin een representatie het realiteitsgehalte van het reële object al of niet benaderd.

Er kan worden gevarieerd door de hoeveelheid details die in elk (deel)object wordt weergegeven. In het hier gepresenteerde onderzoek wordt dit 'realisme' genoemd.

Variatie is er ook in het realiteitsgehalte van een representatie door (deel)objecten afzonderlijk van elkaar of binnen het geheel van onderlinge relaties en samenhang te tonen. In het hier gepresenteerde onderzoek wordt dit als 'context' benoemd.

In de volgende subparagrafen worden beide kenmerken verder onderzocht.

Effect van realisme in illustraties op instructie en toetsing

Diverse auteurs (Anglin, Vaez & Cunningham, 2004; Braden, 1996; Levie, 1987) die onderzoeken naar de effectiviteit van illustraties aan teksten hebben geanalyseerd of zelf zulk onderzoek hebben verricht (Dwyer, 1978), kwamen tot de al in hoofdstuk 1 genoemde conclusie dat de mate van realisme in een illustratie relevant is. Over wat precies de invloed is op het leerproces als er variaties zijn in de mate waarin een illustratie een reëel object weerspiegelt zijn de onderzoeksresultaten verdeeld. Zo geeft Brody (1982) aan dat foto's en gedetailleerde realistisch ogende tekeningen in principe de voorkeur hebben van de lerenden in vergelijking tot simpelere visualisaties zoals schematische tekeningen. Maar deze voorkeur blijkt niet gerelateerd te zijn aan hun uiteindelijke leerprestatie.

Loftus en Bell (1975) hebben onderzoek verricht naar de herkenbaarheid van plaatjes waarbij de vraag was in hoeverre herkenning plaatsvond op basis van globale informatie zoals contouren in een tekening of op basis van informatieve details die aanwezig zijn in foto's. Het experimenteel materiaal bestond uit twee sets van 60 drietallen plaatjes. Elk drietal bestond uit een foto van een natuurlandschap, een representatieve contourtekening met daarin de belangrijke detailinformatie in fotovorm en een verfraaide contourtekening waaraan allerlei informatieve details in fotovorm waren toegevoegd. Twee groepen van 90 studenten kregen elk een set foto's te zien en werden daarna getest op herkenning met een set van 120 plaatjes waarvan 60 originele foto's en 60 afleiders. De proefpersonen moesten drie soorten vragen beantwoorden, namelijk a) of ze het plaatje herkenden, b) of ze dat zeker wisten en c) aangeven wat het plaatje voorstelt door het te benoemen of te beschrijven. De metingen werden viermaal herhaald en uit de resultaten bleek dat foto's beter werden herkend dan tekeningen maar dat er geen verschil is in herkenning tussen de twee tekenvormen in dezelfde set. Volgens de onderzoekers werd dit resultaat veroorzaakt door de mate van details want foto's bevatten veel meer informatieve details dan tekeningen. Deze testresultaten zouden het bewijs zijn dat herkenning plaatsvindt op basis van detailinformatie. In soortgelijk onderzoek vonden Borges, Stepnowsky en Holt (1977) dezelfde resultaten, zoals eerder vermeld in hoofdstuk 1. Herkenning van kleurenfoto's bleek significant hoger dan herkenning van zwart-wit foto's en herkenning van deze twee representatievormen significant hoger dan herkenning van woorden. Het onderzoek werd uitgevoerd met studenten die dertig onderwerpen op dia's te zien kregen. Deze dertig onderwerpen werden getoond in drie verschillende

presentatievormen door elke tien onderwerpen in één specifieke presentatievorm te tonen. Deze presentatievormen gaven het object tekstueel weer in een woord dan wel in een zwart-wit foto dan wel in een kleurenfoto.

In tegenstelling tot deze conclusies bewijst onderzoek van Nelson, Metzler en Reed (1974) juist dat er weinig significant verschil in herkenning is tussen foto's en tekeningen, zelfs na zeven weken. Deze onderzoekers vergeleken de herkenning van gedetailleerde kleurenfoto's met die van verfraaide tekeningen, van simpele lijntekeningen en van woorden. De drie typen representatieve illustraties werden alle drie beter herkend dan woorden maar onderling was er geen significant verschil. De mate van details verwerkt in de representatie maakte dus geen verschil en zelfs bij herhaling van de test na zeven weken werd er geen verschil gevonden. Wel was er sprake van een plafondeffect in herkenning als de test direct na het experiment werd uitgevoerd. Dit laatste wordt bevestigd door Levie (1987) die aangeeft dat bij visueel leren lerenden in veel voorkomende gevallen hoog scoren op herkenningstaken. In het ontwerpen van een toets voor het hier gepresenteerde onderzoek, hoofdstuk 4, wordt dit aspect meegenomen.

Voor het in dit onderzoek gepresenteerde toetsingsprobleem zouden, volgens Loftus en Bell, foto's beter zijn voor het herkennen dan tekeningen terwijl Nelson, Metzler en Reed aangeven dat het niet uitmaakt of foto's dan wel tekeningen worden gebruikt. Ritchey, (1982) deed bij zowel kinderen als volwassenen soortgelijk onderzoek waarbij een vergelijking is gemaakt tussen het herkennen van specifieke objecten in gedetailleerde tekeningen, in simpele lijntekeningen in de vorm van contourtekeningen en in woorden. Aan dit onderzoek deden 144 schoolkinderen en volwassenen mee. Het onderzoeksmateriaal bestond uit een set van 54 items met drie formats waarvan a) enkel de naam van het object, b) een contourtekening met de naam en c) een gedetailleerde tekening met de naam. De personen kregen 18 items te zien waarna ze getoetst werden op herkenning met alle items. Uit dit onderzoek blijkt dat contourtekeningen beter werden herinnerd dan gedetailleerde tekeningen en beide weer veel beter dan woorden. Dit leidde tot de conclusie dat het toevoegen van meer detail niet leidde tot het beter herinneren van de plaatjes.

Deze onderzoeken hebben de gemeenschappelijke conclusie dat toevoeging van visuele representaties bijdraagt aan het herkennen van een object boven het enkel aanbieden van een woord. Zij bieden geen uitsluitsel over de hoeveelheid realistische

detaillering die in een illustratie minimaal vereist is voor instructie of toetsing. Vanuit de perceptietheorie van Arnheim en de CTML theorie wordt de suggestie gewekt dat schematische illustraties voldoende zijn voor het herkennen terwijl de perceptietheorie van Gibson juist zo realistisch mogelijke illustraties suggereert.

Deze beschreven onderzoeken bekrachtigen de eerdere aangegeven algemene conclusie (Anglin, Vaez & Cunningham, 2004; Braden, 1996; Levie, 1987) dat realisme een belangrijk aspect is in onderzoek naar visuele representaties maar dat het niet mogelijk is om eenduidig aan te geven of uit te sluiten welke kenmerken/eigenschappen van realistische illustraties het beste passen bij de gestelde leerdoelen voor het verkrijgen van het construct of het toetsen ervan.

Voor het hier gepresenteerde onderzoek is het volgende beredeneerd. Het is een gegeven dat studenten leren aan de hand van het reële object, de rat. De leerdoelen in de instructie zijn gericht op het leren van de statische configuratie van de rat aan de hand van diverse dissectielagen waarbij een aantal lagen een fysiek systeem weergeven. Het te leren construct is dan ook de statische configuratie van de diverse anatomielagen van de rat zoals de verschillende delen, de samenstelling van deze delen, hoe ze verbonden zijn en ruimtelijk geconfigureerd. Zowel foto's als schematische tekeningen kunnen hierin effectief bijdragen aan het leerproces waarbij uiteindelijk moet worden getoetst of de lerende het beoogde construct beheerst. De literatuur biedt hierin geen duidelijkheid over de relatie tussen realisme en het toetsen van het construct maar enkel aanwijzingen ten aanzien van de rol van realisme in het leerproces. Dit roept de vraag op welke verschijningsvorm, variaties op realisme, het meest bijdraagt aan het toetsen van dat construct. Door realisme te onderzoeken vanuit een gedetailleerde weergave zoals een foto enerzijds versus een gesimplificeerde weergave zoals een schema anderzijds kan dat worden onderzocht. Dit resulteerde in de volgende onderzoeksvraag.

Meet een toets met foto's hetzelfde visuele construct als een toets met schema's?

Deze specifieke onderzoeksvraag zal nader worden onderzocht in experimenten die worden beschreven in de hoofdstukken 5 en 6.

Effect van context in illustraties op instructie en toetsing

Figuur 2.1 alsook tabel 2.1 duiden op twee manieren aan hoe kan worden omgegaan met de weergave van de realiteit in een illustratie. Realistisch in termen

van fotovorm en schemavorm waarbij het gehele beeld met (deel)objecten behouden blijft of realistisch in termen van het aanbieden van het geheel met alle (deel)objecten of één enkel (deel)object. In beide gevallen is het de vraag hoeveel visuele informatie er in een representatie nodig is om het geleerde construct te toetsen en tegelijkertijd nog een zinvolle weergave te zijn van het referentieobject. In het hier gepresenteerde onderzoek is het construct gericht op de statische configuratie van de anatomie van de rat waarbij anatomische objecten in de rat zich op een bepaalde wijze tot elkaar verhouden door ondermeer vorm, positie, grootte, textuur en andere kenmerken die gezamenlijk een Gestalt vormen. Volgens de Gestaltprincipes (Arnheim, 1974) is perceptie gericht op het begrijpen van definieerbare vormen en patronen. De Gestalt wordt gevormd door de groepering van stimuli tot één geheel, dat in de waarneming centraal staat. Kenmerken van een Gestalt zijn: (a) dat het geheel eerder wordt waargenomen dan de delen waaruit de Gestalt is opgebouwd, (b) dat als één element binnen Gestalt verandert, de Gestalt meteen als totaliteit verandert en (c) dat de Gestalt altijd meer is dan de som van de aparte elementen waaruit deze is samengesteld. In de literatuur is, in de lijn van de hier gepresenteerde onderzoekssituatie, geen vergelijkbaar onderzoek gevonden naar de effectiviteit van visuele representaties als de Gestalt in een representatie onvolledig is of het referentieobject niet geheel weergeeft. Een voorbeeld daarvan is als deelobjecten in een representatie los van elkaar worden weergegeven zonder weergave van de begrenzingcontouren van het referentieobject zoals in het hier gepresenteerde onderzoek de anatomische structuren van de spijsvertering van de rat worden weergegeven zonder de contouren van de rat. Is de maag, figuur 2.1 versie C & D, dan een weergave van een rattenmaag of de weergave van een maag van een ander zoogdier of van een andere klasse dieren?

Hoewel binnen de visuele geletterdheid en visueel leren geen onderzoeken beschikbaar waren die dit aspect onderzochten, blijkt binnen de (neuro)psychologie (Tanaka & Farah, 1993; Moscovitch & Moscovitch, 2000; Cave & Kosslyn, 1993) wel aandacht te worden besteed aan dit aspect met betrekking tot onderzoek naar het herkennen van gezichten.

In hoofdstuk 1 zijn de relevante onderzoeken (Moscovitch & Moscovitch, 2000; Palmer, 1975) kort beschreven. Moscovitch en Moscovitch (2000) deden hun onderzoek naar de identificatie van gezichten met (a) beelden van gezichten met contextuele informatie zoals de gezichtscontouren, (b) beelden van gezichten die op de kop stonden en (c) beelden waarvan de gezichtsdelen in nieuwe ruimtelijke relaties ten opzichte van elkaar stonden waardoor het beeld verstoord was. Uit

de resultaten bleek dat gezichten met de juiste ruimtelijke relaties en in de correcte contouren het best werden herkend en dat de herkenning daalde naarmate het beeld meer werd verstoord. Wel was herkenning van afzonderlijke onderdelen altijd hoog. Deze bevindingen zouden ook gelden voor het herkennen van andere objecten dan gezichten. Palmer (1975) heeft hetzelfde onderzoeksprincipe toegepast in de context van de keuken en objecten die daaraan gerelateerd zijn. Palmer (1975) gebruikte objecten die voornamelijk te vinden zijn in een keuken die (a) in een conditie met goede context werden getoond in een keukenomgeving, (b) in een conditie zonder context afzonderlijk werden getoond en (c) in een conditie met verkeerde context werden getoond in een andere omgeving dan de keuken. Objecten werden het best herkend in aanwezigheid van de correcte contextuele informatie en het slechtst werden objecten herkend in aanwezigheid van de verkeerde contextuele informatie. Deze resultaten kunnen worden verklaard vanuit het uitgangspunt van de Gestalt-regels (Behrmann & Moscovitch, 2001; Moscovitch & Moscovitch, 2000) die aangeven dat herkenning afhankelijk is van de organisatie van informatie naar aanleiding van bepaalde principes. De objecten zijn van belang maar vooral de organisatie van de objecten ten opzichte van elkaar ofwel de ruimtelijke relaties tussen de objecten zijn bepalend. In het geval van gezichtsherkenning zijn er vast omschreven regels over de organisatie van de afzonderlijke gezichtsdelen of objecten. De organisatie van deze objecten vormt samen een vast patroon dat resulteert in een bepaalde context. Wordt dit patroon verstoord dan wordt het herkennen van het geheel moeilijker. Een neus wordt wel als neus herkend en een oog wel als oog maar herkenning van het gezicht is moeilijker.

In het hier gepresenteerde onderzoek leert de lerende in de instructie alle anatomische onderdelen binnen de context van het reële object, de rat. Dat betekent dat het leerproces er ook op is gericht dat alle afzonderlijke anatomische onderdelen van de rat in samenhang, in een Gestalt, worden geleerd. Van de lerende wordt verwacht dat deze de afzonderlijke anatomische onderdelen zowel afzonderlijk als in samenhang (topografisch) (her)kent. Voor het toetsen van het geleerde construct is niet duidelijk hoe noodzakelijk het is te toetsen met of zonder context. Zijn de afzonderlijke anatomische objecten binnen in de rat ook geïsoleerd herkenbaar of moet dat altijd in relatie tot de contouren van de rat of daaraan gerelateerde anatomische objecten? Dit vraagstuk resulteert in de tweede specifieke onderzoeksvraag:

Meet een toets met context hetzelfde visuele construct als een toets zonder context?

Ook deze specifieke onderzoeksvraag zal nader worden onderzocht in experimenten die worden beschreven in de hoofdstukken 5 en 6.

2.4 VISUELE REPRESENTATIES IN RELATIE TOT VISUEEL TOETSEN

Het hier gepresenteerde onderzoek is gestart vanuit het in hoofdstuk 1 beschreven toetsprobleem dat zich voordoet binnen het biologieonderwijs. Dit probleem dient te worden opgelost door het ontwikkelen van een toets die (a) aansluit op het leerproces, (b) objectief is en (c) de studenten in een relatief kort tijdsbestek van feedback voorziet. Tot nu is in dit hoofdstuk aandacht besteed aan het construct, de leerprocessen die zich daarbij voltrekken in relatie tot de visuele variabelen die daarin een rol spelen en daarvoor relevante instructiemethoden. In deze paragraaf wordt ingegaan op de toetsvormen die de resultaten van het visueel leerproces, het gevormde construct, kunnen toetsen in relatie tot de gestelde leerdoelstellingen.

Visueel leren betekent visueel toetsen

Braden (1996) stelt dat voor wat betreft het toetsen van kennis opgedaan door visueel leren, zogenaamde visueel toetsen, weinig onderzoek voorradig is in vergelijking tot onderzoek naar visueel leren. De meeste onderzoeken met visuele toetsen zijn verricht in het PSE-programma van Dwyer en zijn medewerkers. De algemene conclusie (Braden, 1996; Moore & Dwyer, 1994) is dat kennis opgedaan in een leeromgeving met visuele representaties, visueel leren, het best getoetst kan worden met visuele representaties, visueel toetsen. Naar aanleiding van zijn studies rapporteert Dwyer (1994) dat de combinatie van visuele instructie en visuele toetsing de prestatie bevordert en dat visueel toetsen een valide strategie is voor het toetsen van kennis opgedaan met een visuele instructie. Dwyer (1978) geeft als belangrijke motivatie voor het ontwikkelen van visuele toetsen aan dat de combinatie van visuele instructie en visueel toetsen de score van de student verbetert omdat de zintuigmodaliteit gehandhaafd blijft wat vervolgens ook de betrouwbaarheid van het toetssysteem ten goede komt. Hij stelt dat door de grote mate van overeenkomst tussen de onderwijsleersituatie (coderingsfase) en de toetssituatie (retrieval fase) visueel toetsen effectief is. De toetssituatie is daarmee een weerspiegeling van de leersituatie. Een voorbeeld dat Dwyer ter bewijsvoering aanhaalt, is het experiment van Dwyer en De Melo (1983) waarbij

visuele toetsen, zijnde tekstuele toetsen met visualiteit, werden vergeleken met uitsluitend tekstuele toetsen voor het meten van een visuele instructie. De toegevoegde visualiteit bestond uit een aantal plaatjes. Uit de resultaten bleek dat de prestaties op visuele toetsen hoger waren dan op de tekstuele toetsen. Dit resultaat is niet verbazingwekkend want de onderzoeksopzet omvat niet meer dan een vergelijking tussen een tekstuele instructie zonder visuele ondersteuning en een tekstuele instructie met visuele ondersteuning. De onderzoeksvraag behelst dan in hoeverre visuele ondersteuning aan tekstuele toetsen een toegevoegde waarde heeft.

Toch wordt in het hier gepresenteerde onderzoek gesteld dat het gedefinieerde construct het best kan worden getoetst met een visuele toets. Dat is niet zozeer gebaseerd op de onderzoeksresultaten zoals het hier beschreven onderzoek van Dwyer en De Melo (1983) maar vooral op de redenering die Dwyer (1978) hiertoe aanhoudt, het handhaven van de modaliteit waardoor leersituatie en de toetssituatie op elkaar zijn afgestemd. Dit kan worden beargumenteerd op basis van het principe van actief leren uit de CTML theorie (Mayer, 2005, beschreven op blz. 23 en 24), waarbij de lerende cognitieve processen toepast op inkomende informatie om zo zinvolle informatie-eenheden te genereren met als resultaat de constructie van een coherente mentale representatie. Door te toetsen met dezelfde modaliteit als waarin wordt geleerd, mag worden aangenomen dat deze de zinvolle informatie-eenheden sneller heractiveert waarmee de juiste mentale representatie(s) gemakkelijker wordt(en) opgehaald. Is de instructie visueel dan is een toets met visuele representaties zinvoller dan een toets met alleen tekstuele vragen.

Visueel toetsen: leerdoelen en toetstaken

In Dwyer's PSE-programma is een toetsmethode voor visueel leren ontwikkeld die aanknopingspunten biedt voor het hier gepresenteerde onderzoek (Dwyer, 1978, 1994; Braden, 1996; Anglin, Vaez & Cunningham, 2004). Zoals eerder aangegeven is de toets vooral gericht op het toetsen van de effectiviteit van verbale instructies, tekstueel dan wel mondeling, die worden ondersteund door variërende visuele representaties. In het PSE-programma krijgt dat als volgt vorm. Het programma bestaat uit een reeks van gerelateerde experimenten tijdens de periode 1978 tot 2007 met als onderwerp de anatomie van het kalfshart. Elk experiment behelst het vergelijken van twee verbale instructies waarvan één wordt ondersteund met visuele representaties. Op basis van de experimentele vraag zijn de onafhankelijke variabelen variaties in leerling-kenmerken, gebruikte

instructiemedia en toegepaste visualisaties. Om de experimentele condities te toetsen is een meerdelige toets ontwikkeld (de PSE-toets) waarmee steeds getoetst is welke visualisatievormen in een instructie onder bepaalde condities het effectiefst zijn. De PSE-toets is gedurende het onderzoeksprogramma onderzocht, doorontwikkeld en waar nodig bijgesteld.

De PSE-toets definieert het construct voor de anatomie van het kalfshart. De toets is onderverdeeld in deeltoetsen voor de verschillende leerdoelen van de instructie. Daarbij ging Dwyer (1978) uit van het bestaan van verschillende typen kennis die hij relateerde aan de leertaken (leerdoelen) en vervolgens aan toetstaken. Zoals eerder aangegeven zouden deze kennistypen hiërarchisch geordend zijn van simpel naar complex. De PSE-toets bestaat daarmee uit vier deeltoetsen die samen bepalen of een student geslaagd is voor dat studieonderdeel. Elke deeltoets bestaat uit twintig items en meet het bereiken van een eigen onderwijsdoel. De deeltoetsen zijn (1) de terminologietoets, (2) de identificatietoets, (3) de tekentoets en (4) de begripstoets. De terminologietoets toetst de mate waarin de student kennis heeft van feiten en definities. De identificatietoets toetst de mate waarin de student in staat is delen van het hart of posities van een object binnenin het hart te identificeren. De tekentoets toetst in hoeverre een student een object kan reproduceren in de correcte context. De begripstoets toetst in hoeverre een student complexe procedures en/of processen beheerst zoals de relatie van een object tot andere objecten afhankelijk van functie en positie.

Zoals eerder omschreven zijn voor het hier gepresenteerde onderzoek de leerdoelen in overeenstemming met het te toetsen construct: (1) het kunnen herkennen van visuele kenmerken op basis van vorm, kleur en textuur, (2) het kunnen benoemen van de visuele kenmerken en (3) het kunnen aangeven hoe de structuren ruimtelijk aan elkaar zijn gerelateerd rekening houdend met de relatieve groottes ten opzichte van elkaar. Voor het hier beschreven onderzoek kunnen net als in de PSE-toets de leerdoelen voor de hier te ontwerpen toets vertaald worden naar drie deeltoetsen die inhoudelijk dezelfde onderwerpen toetsen. Er ontstaat dan een toets bestaande uit drie deeltoetsen waarvan een eerste identificatietoets met als taak het herkennen van visuele kenmerken van anatomisch objecten, een tweede identificatietoets met als taak het benoemen van de visuele kenmerken met de Latijnse naam en tot slot een topografische toets met als taak het aangeven hoe structuren ruimtelijk aan elkaar zijn gerelateerd, conform de gestelde leerdoelen.

In het onderzoeksprogramma met de PSE-toets is Dwyer (1978) op basis van werk van Gagné (1965) uitgegaan van een hiërarchie in de leerdoelen die loopt van simpel naar complex. Deze leerdoelen koppelt Dwyer vervolgens aan de toetstaken en beredeneert dat deze elk een weergave zijn van een bepaald beheersingsniveau. Uit zijn werk wordt niet duidelijk wat de verschillen in moeilijkheidsgraad veroorzaakt. Kunnen de verschillen in moeilijkheidsgraad tussen de toetstaken worden herleid naar de hiërarchische opbouw van de leerdoelen? Of worden de verschillen in moeilijkheidsgraad veroorzaakt door de complexiteit van de toetstaken zelf zoals type vraag, type respons of toegepaste visualisatie? Zoals eerder aangegeven onderscheidt ook Hegarty (2005) kennistypen maar zij geeft geen uitsluitsel over een mogelijke hiërarchische ordening.

Toetsen met tekeningen in de biologie

Naast visuele toetsen zoals de PSE-toets van Dwyer (1994) wordt binnen de biologie de nauwe observatie van biologische structuren vooral getoetst door het nakijken van door de studenten geproduceerde tekeningen. Door de productie van tekeningen zouden studenten kennis krijgen over het morfologisch belang van de architectuur van het anatomisch object (Gaupp, 1904). Volgens Jones, Reed en Weyers (1994, 1998) heeft het maken van een tekening allereerst een instructiedoelstelling namelijk het opdoen van observatievaardigheden. Eerder (blz. 49) is beredeneerd dat tekenen als taakgerichte opdracht sturing geeft aan het leerproces en als zodanig inderdaad bijdraagt aan het vormen van een correct mentaal model. Jones, Reed en Weyers (1994, 1998) gaan vervolgens een stap verder en leggen een relatie tussen de kwaliteit van de tekening en het te leren construct. Een kwalitatief goede tekening getuigt van voldoende kennis over het te leren construct en een kwalitatief slechte tekening van onvoldoende kennis over het visuele construct. Deze benadering veronderstelt dat het eindresultaat van de tekentaak, de externe visuele representatie, een één op één weergave is van het gevormde mentaal model. Of Jones, Reed en Weyers daarin gelijk hebben, kan niet met zekerheid worden gezegd. Er kan niet worden uitgesloten dat een goede tekening mogelijk meer getuigt van de goede tekenvaardigheden van de lerende en in mindere mate een maat is voor het hebben verworven van een correct mentaal model. Deze toetsvorm is niettemin van oudsher in de biologie in gebruik en neemt nog steeds een voorname plek in.

Op basis van de veronderstelde validiteit van toetsen naar aanleiding van door de lerenden gemaakte tekeningen is voor het practicum dat in dit onderzoek centraal

staat een prototype van een digitale toets ontwikkeld om deze tekeningen te corrigeren, genaamd Ratview (De Graaff & Verhagen, 1997). De door de studenten tijdens het dissectiepracticum gemaakte tekeningen representeren de geobserveerde anatomische structuren. Die structuren zijn door de studenten benoemd door de Latijnse namen van de verschillende componenten er bij te schrijven. Elke student verricht zelf de correctie van de tekening met behulp van een computerprogramma. De student gaat daarvoor na het maken van een tekening naar de computer en tracht met de tekening erbij het geobserveerde beeld opnieuw te construeren. Eerst voert de student alle Latijnse namen van de anatomische objecten in, daarna maakt de student een keuze uit een verzameling grafische representaties van verschillende anatomische objecten (waaronder ook afleiders) en koppelt elk gekozen anatomisch object aan de bijbehorende naam. In de laatste fase construeert de student zijn tekening met behulp van de zo verkregen bouwstenen binnen de op het beeldscherm getoonde contouren van een rat. Na de reconstructie bepaalt het programma in hoeverre de student een correcte reconstructie heeft gemaakt. De resultaten van de beoordeling van de tekening geconstrueerd op de computer werden vervolgens vergeleken met de resultaten van handmatig geproduceerde tekening. Het cijfer toegekend door de docent aan de handmatig geproduceerde tekening en het cijfer toegekend door het computerprogramma aan de gereconstrueerde tekening verschilden significant waarbij de cijfers toegekend door het toetsprogramma consistent lager waren. De verklaring hiervoor is dat het programma op eenduidige wijze de tekeningen beoordeelde terwijl bij de handmatige beoordeling de docenten meer aspecten afwegen die bij elkaar genomen resulteerden in een hoger cijfer. Een voorbeeld daarvan is dat het programma niet de netheid van een geproduceerde tekening afweegt in de eindbeoordeling terwijl een docent dat impliciet wel in de beoordeling mee neemt wat kan leiden tot een hoger cijfer. De toets kwam daarmee niet geheel overeen met de beoordeling van de docent. Niettemin voldeed de toets op verschillende punten aan de eisen voor het oplossen van het toetsingsprobleem omdat de student onmiddellijk feedback kreeg over de geproduceerde tekening, de toetsingsmethodiek gestandaardiseerd en objectief was en aansloot op het leerproces. Echter de vraag die hier moet worden gesteld, is of de toets dekkend is voor het toetsen van het geleerde construct. Toetst het programma alleen of de student een goede tekening heeft gemaakt en daarmee de instructieopdrachten goed heeft vervuld, of toetst het programma inderdaad het geleerde construct? Bij het ontwerpen en construeren van de toets in dit onderzoek is dit een richtinggevende vraag. Met het te verwerven visuele construct als uitgangspunt, is het noodzakelijk de toetsingsmogelijkheden voor visueel toetsen verder te verkennen.

Toetsmogelijkheden voor visueel toetsen

In de loop der tijd zijn, mede door technologische ontwikkelingen, de toepassingen voor het toetsen met visuele representaties toegenomen. Parshall, Davey, and Pashley (2000) hebben getracht de diversiteit aan toepassingen te classificeren wat resulteerde in vijf dimensies, te weten (a) toetsstructuur, (b) soort response, (c) multimediatoepassingen, (d) mate van interactiviteit en (e) scoringsmethodiek.

Toetsstructuur is gericht op de presentatie van de toetsvorm en de daarbij verwachte reactie van de toetskandidaat. Dat varieert van selectie van het correcte antwoordalternatief, zoals meerkeuze-items, tot diverse mogelijkheden voor de constructie van het antwoord door de toetskandidaat, zoals open vragen. Items gebaseerd op selectie bevatten een vraagstelling met verschillende antwoordalternatieven waarbij de toetskandidaat voor een of meer alternatieven kan kiezen. Voor visuele toetsen kunnen figuren en diagrammen in dit type itemformat worden gebruikt zowel in de stam als in de alternatieven (O'Neill & Folk, 1996; Martinez & Katz, 1992). Als het gaat om de categorie items met constructietaken is het mogelijk om deze in een geautomatiseerde toetsvorm te geven met uiteenlopende visuele representaties. Vooral Martinez (1998) heeft zich verdiept in de zo ontstane "constructed figural response" items (CFR-items). Dit zijn items waarbij zoveel mogelijk gebruik wordt gemaakt van visualiteit, voornamelijk met plaatjes (2D en 3D), en de toetskandidaat een constructietaak voorgeschoteld krijgt. CFR-items verschillen van meerkeuze-items op twee punten namelijk (a) CFR-items behelzen een constructietaak in plaats van een selectietaak en (b) CFR-items vereisen van een toetskandidaat dat deze om kan gaan met visueel materiaal (Martinez, 1990). Itemformats met een reactie op basis van selectie zoals meerkeuze worden vaak toegepast in schriftelijke en vooral geautomatiseerde toetsen omdat ze, in vergelijking tot constructie-items: (a) een relatief korte testtijd vereisen, (b) gemakkelijker te scoren zijn, (c) goedkoper in productie zijn en (d) een hoge betrouwbaarheid in testscore hebben (Downing, 2006; Sireci & Zenisky, 2006). Downing (2006) geeft wel aan dat CFR-items meer mogelijkheden bieden voor het toetsen van meer complexe kennis en vaardigheden. Om die reden is het de moeite waard CFR-items te overwegen in plaats van meerkeuze-items. In de volgende paragraaf worden op dit aspect CFR-items en meerkeuze-items vergeleken.

De tweede dimensie, soort respons, omvat de verschillende responsvormen die worden getypeerd door de fysieke acties die een toetskandidaat moet verrichten

om een item te kunnen beantwoorden. Voor pen-en-papier toetsen ligt de responsvorm voor de hand. Voor geautomatiseerde toetsen zijn het gebruik van een toetsenbord en een muis voorbeelden van responsvormen. Voor een geautomatiseerde visuele toets biedt het gebruik van de muis verschillende mogelijkheden zoals het aanklikken van een object om deze vervolgens te verslepen naar een ander object of met een tweede klik te verbinden met een ander object (Martinez & Katz, 1992; Bejar & Braun, 1999).

Het gebruik van beelden, geluid en video zijn voorbeelden van de derde dimensie, multimediatoepassingen. Het biedt de mogelijkheid tot het gebruik van visualisaties waarmee toetskandidaten diverse handelingen kunnen uitvoeren zoals roteren, van formaat veranderen en in- en uitzoomen. Toetskandidaten kunnen plaatjes verslepen en (re)constructietaken uitvoeren (Martinez, 1998). De vierde dimensie, de mate van interactiviteit, omvat de mate waarin het toetsprogramma reageert op de input van een toetskandidaat. Deze interactie maakt onder andere adaptief toetsen mogelijk, het inspelen op het toetsgedrag van de toetskandidaat tijdens het toetsproces.

Tot slot is er de vijfde dimensie, de scoringsmethode. Door de uiteenlopende mogelijkheden binnen deze vijf dimensies is er een enorme diversiteit aan toetsen ontstaan waarbij toetsen steeds complexer zijn in de samenstelling daarvan. Complexere samengestelde toetsen vereisen vervolgens steeds complexere scoringsmethoden om tot een objectieve score te komen. Verschillende onderzoeken (Bejar, 1991; Bennett & Bejar, 1998) naar geschikte scoringsmethoden geven dan ook aan dat de complexiteit toeneemt naarmate de toets ingewikkelder is.

De classificatie in vijf dimensies is actueel maar eveneens verfijnd door de toevoeging van nog twee dimensies: complexiteit en precisie (Parshall, Harmes, Davey & Pashley, 2010). De dimensie complexiteit richt zich op het aantal en de variatie van elementen die toetskandidaten moeten interpreteren en gebruiken om aan de eisen van het toetsitem te kunnen voldoen. De dimensie precisie is gericht op de mate waarin een item een realistische en accurate representatie is van de feitelijke objecten, situaties en taken die onderdeel zijn van het te toetsen construct. Onderzoekers geven aan dat toename in precisie kan leiden tot dure programma's. Voor het hier gepresenteerde onderzoek is in het vorige hoofdstuk aangestipt dat het te ontwikkelen toetsprogramma 'productief' dient te zijn, ofwel een afweging tussen de effectiviteit (te bereiken resultaat) en de efficiëntie (relatie kosten/baten) van het programma is hierin nodig. Tijdens de toetsontwikkeling

zal steeds de balans moeten worden opgemaakt tussen het vereiste precisieniveau in de toets, zoals gedetailleerdheid representaties, en de kosten die daarmee gemoeid zijn. Dat onderschrijft onder meer ook het belang van onderzoek naar het vereiste realiteitsgehalte in de representaties (realisme en context). Gedetailleerdere representaties zijn in veel gevallen ook duurder.

Meerkeuze items versus CFR-items

Martinez (1998) en Martinez en Katz (1992) ontwikkelden verschillende CFR-items met diverse responsevormen zoals het verslepen, het verbinden en modelleren van objecten.

Tevens onderzochten zij (Martinez & Katz, 1992) de verschillen tussen CFR-items en meerkeuze-items in het domein van de architectuur. In hun onderzoek worden drie categorieën van elkaar onderscheiden op basis van de verschillende te toetsen taken. Deze categorieën zijn (a) declaratieve kennis waarin op feitenkennis wordt getoetst, (b) procedurele kennis waarin toetskandidaten een algoritme toepassen voor het beantwoorden van een vraag en (c) 'discovered strategy items' waarin toetskandidaten een oplossing moeten bedenken voor een item in puzzelvorm. Van deze drie categorieën zijn zowel CFR-items als meerkeuze-items getoetst. De onderzoeksresultaten geven aan dat van de items die betrekking hadden op dezelfde kennis, de CFR versie moeilijker was dan de meerkeuzeversie. De onderzoekers geven als mogelijke verklaring dat dit veroorzaakt wordt door de verschillende oplossingsstrategieën die worden toegepast op meerkeuze-items en CFR-items. Hier wordt in de literatuur verder niet specifiek op ingegaan. Vervolgens bleek ook dat de drie CFR-itemformats van elkaar verschilden in moeilijkheidsgraad waarbij 'discovered strategy items' het moeilijkst zijn. Dit zou volgens hen kunnen wijzen op toetstaken die mogelijk hiërarchisch geordend zijn. De onderzoekers hebben daar verder geen expliciete uitspraken over gedaan. Mogelijk zit de verklaring in de taxonomie die Martinez (1998) heeft gemaakt van een scala aan CFR-items dat is gerelateerd aan de cognitieve processen die vereist zijn voor het oplossen ervan.

De taxonomie van Martinez (Tabel 2.2) onderscheidt zeven taken en koppelt aan elke taak het gegeven waarop de taak moet worden toegepast, de bijbehorende cognitieve processen en de scoringscriteria.

Tabel 2.2 *Taxonomie CFR-items (Martinez, 1998)*

Type	Task	Givens	Cognitive processes	Scoring criteria
1	Point to image element	Name of element and complex image	Recognize/recall image; possibly isolate image element perceptually	Location within response field
2	Evaluate image	Image	Retrieve name, symbol; compute value/attribute; detect error	Location within response field; selection/construction of evaluation
3	Show direction/path of motion	Initial trajectory, vectors, path criteria	Determine direction or path	Match with ideal direction or path
4	Show height, extent or boundaries	Referenced background	Determine how much, where, and in what position	Location within response field or position relative to other marks
5	Assemble elements; possibly create elements	Element set and/or a partial structure	Determine structure	Location, orientation
6	Indicate categorical, ordinal or functional relationships	Concepts and group names, dimension axes, matrix or links	Determine categorical identities, orders, or relationships	Location, relation
7	Indicate continuous relationships	Data or concepts and dimension axes	Determine continuous relationships linear or non-linear	Location, orientation, straight or relation free-form

In tabel 2.3 is deze taxonomie toegepast op de typen leerdoelen voor het hier gepresenteerde onderzoek. Dit levert een overzicht op van itemformats die voor het hier gepresenteerde onderzoek relevant kunnen zijn.

Tabel 2.3 *Relatie leerdoeltype en toetstaken*

Leerdoelen n.a.v. het construct	Toetstaken Mogelijke typen CFR items (conform tabel 2.2)
Benoemen visuele kenmerken	2
Herkennen van visuele kenmerken naar vorm, kleur en textuur	2 & 4
Kunnen aangeven hoe de structuren ruimtelijk aan elkaar gerelateerd zijn	3 & 5 & 6

In de hoofdstukken 3 en 4 bij de ontwikkeling van de toets wordt dit nader uitgewerkt in de toetstaken.

2.5 TE ONDERZOEKEN VRAAGSTELLINGEN NAAR AANLEIDING VAN DE LITERATUUR

Aan het begin van dit hoofdstuk werd gesteld dat de informatie uit de gepresenteerde literatuur een aantal argumenten oplevert om (a) instructie en toets in te richten voor het bereiken van de onderwijsdoelstellingen, (b) de geformuleerde onderzoeksvraag te vertalen naar specifieke onderzoeksvragen en (c) variabelen te definiëren voor de experimenten ter beantwoording van die meer specifieke onderzoeksvragen.

De algemene onderzoeksvraag van waaruit is vertrokken luidt:

‘Wat is een effectieve en efficiënte combinatie van instructie en toetsing voor het bereiken van onderwijsdoelstellingen die verbonden zijn met leren door observeren van concreet biologisch materiaal’.

De literatuur heeft daarvoor het volgende opgeleverd:

1. Relevantie van taakgerichte opdrachten bij sturing van selectie van visuele stimuli

De waarnemer heeft een actieve rol bij de interpretatie van een plaatje door diens verwachtingen, diens doelstelling (exploratief of taakgericht) en diens kenmerken (leeftijd, oogsterkte, cultuur, etc). Het leren van plaatjes in het visueel leerproces vereist doelbewuste aandacht en het geven van sturing in de vorm van taakgerichte opdrachten draagt daar aan bij. Onderzoeken naar taakgerichte opdrachten hebben als resultaat opgeleverd dat ze inderdaad de aandacht sturen, sturend zijn ten aanzien van die visuele kenmerken die van belang zijn en als zodanig kunnen bijdragen aan het bereiken van de gestelde leerdoelen in de instructie, het te leren construct. Tekenen is een vorm van een taakgerichte opdracht maar er zijn ook alternatieven zoals het benoemen van tekeningen van het reële object. Voor het hier gepresenteerde onderzoek is deze informatie relevant omdat een instructie met taakgerichte opdrachten anders dan tekenen mogelijk even effectief is. Dit is van belang ervan uitgaand dat de te construeren toetsmethode de correctie van tekeningen niet meer toelaat waardoor de huidige instructie daardoor niet afdoende meer op de toets aansluit. In hoofdstuk 3 wordt een onderzoek gepresenteerd met als vraagstelling:

Is de vernieuwde instructie gelijkwaardig aan de bestaande instructie voor het bereiken van de onderwijsdoelen van het dissectiepracticum anatomie van de rat?

2. Vereiste hoeveelheid visuele stimuli voor het toetsen van een visueel construct
Literatuuronderzoek naar factoren die mogelijk van invloed zijn bij het toetsen van een visueel construct heeft geleid tot een eerste rangschikking van visualiteitsvormen, tabel 2.1, die zijn gerelateerd aan hun functie en hun bijdrage aan te verkrijgen kennistypen. Daarop voortredenerend blijkt de inbreng van de variabelen realisme en context van belang. De onderzoeksresultaten uit de literatuur zijn tegenstrijdig in de resultaten naar de hoeveelheid realiteit (schematisch vs. realistisch) die noodzakelijk is voor gebruik in visuele representaties. Algemene conclusie is dat realisme een relevant aspect is maar met verschillende effecten. Eenduidige algemene conclusies zijn er onvoldoende.

Ten aanzien van de variabele context liggen de onderzoeksresultaten uit de literatuur op dezelfde lijn namelijk dat het aanbieden van contextuele informatie, zoals contouren rondom objecten die bij elkaar horen, kunnen bijdragen aan herkenning. Onderzoek naar het effect van het al of niet aanbieden van contextuele informatie in een representatie ten behoeve van het leren van een construct of het toetsen ervan is niet voorradig. De beschikbare onderzoeksresultaten zijn afkomstig uit onderzoeken van andere disciplines waaronder onderzoek naar gezichtsherkenning. In het hier beschreven onderzoek wordt verondersteld dat dit ook geldig kan zijn voor het hier gepresenteerde onderzoek.

Op basis van de hier gepresenteerde literatuur zijn de volgende twee specifieke onderzoeksvragen geformuleerd, te weten:

Meet een toets met foto's hetzelfde visuele construct als een toets met schema's?

Meet een toets met context hetzelfde visuele construct als een toets zonder context?

Beantwoording van deze vragen is van belang voor de inrichting van de te construeren toets omdat het eindresultaat bepalend is voor de typen statische visuele representaties die dienen te worden gebruikt.

Deze onderzoeksvragen worden in twee experimenten onderzocht in de hoofdstukken vijf en zes.

3. Typen kennis in het leren van fysieke systemen

In de literatuur is ook ingegaan op het leren van fysieke systemen zoals diagrammen en biologische systemen (Hegarty, 2005). Voor het leren van systemen zou er sprake zijn van verschillende typen kennis, te weten statische

configuratie van het systeem, causale gedrag van het systeem en functies van het systeem. In het hier beschreven onderzoek richt het practicumonderwijs zich vooralsnog op de statische configuratie van het systeem. De theorie biedt argumenten om bij de evaluatie van de instructie en de constructie van de toets stil te staan bij de definitie van het visuele construct en deze te evalueren met als doel na te gaan of deze dekkend is voor de kennis die de lerende opdoet tijdens het leerproces.

4. Toetsmogelijkheden voor visueel toetsen

Voor wat betreft de toetsconstructie stelt de literatuur dat een visuele toets een visueel leerproces beter beoordeelt en de retentie verstevigt door de modaliteit te handhaven dan een toets met een andere modaliteit zoals een tekstuele toets. De mogelijkheden voor visuele toetsen liggen niet alleen meer op het terrein van pen-en-papier-mogelijkheden maar ook binnen de mogelijkheden van geautomatiseerde computertoetsen. Geautomatiseerde toetsen bieden de voordelen van (a) objectieve beoordeling en (b) het in een kort tijdsbestek voorzien in feedback. Deze voordelen lossen het toetsprobleem zoals benoemd deels op. Vervolgens zijn de technologische mogelijkheden met een geautomatiseerde toets dusdanig dat het spectrum aan mogelijkheden voor het construeren van een visuele toets veel ruimer is dan de mogelijkheden die een pen-en-papier-toets biedt. Daarbij kan worden gedacht aan het variëren in itemrespons door het slepen en aanklikken van visuele representaties waardoor onder meer het puzzelen met visuele kenmerken tot de mogelijkheden behoort. Ook het toepassen van multimediatoepassingen en het gebruik van CFR items waarbij toegepaste visuele representaties het referentieobject en de leersituatie zoveel mogelijk kunnen benaderen, bieden voordelen die een pen-en-papier-toets ontbeert. Daarmee zijn er uit de literatuur voldoende argumenten voor het ontwikkelen van een toets met CBT items waarmee de toetsproblematiek wordt verholpen. In de hoofdstuk 3 en 4 wordt de toetsontwikkeling beschreven waarbij de in dit hoofdstuk gepresenteerde informatie bijdraagt aan het ontwikkelen van een CBT-toets die aansluit op de instructie.

5. Hiërarchische ordening in kennissystemen voor instructie en toetsing

Besproken is de mogelijkheid van een hiërarchie in de leertaken naar aanleiding van het PSE-programma. Daarin zijn de leertaken, evenredig aan de leerdoelen, gerelateerd aan een oplopend kennisniveau waarbij gebruik is gemaakt van de principes volgens Gagné. Deze geeft aan dat er een hiërarchische opbouw is in het

verwerven van intelligente vaardigheden met een verloop van simpel naar complex. Deze opbouw is binnen het PSE-programma aangebracht in zowel de leerdoelen als in de instructie als in de toetstaken.

Onderzoek van Martinez naar CFR-items in relatie tot toetstaken geven ook in toetstaken een hiërarchische ordening aan. Voor het hier gepresenteerde onderzoek zijn, weergegeven in tabel 2.3, de leerdoelen van het hier gepresenteerde onderzoek gerelateerd aan de toetstaken die zich lenen voor experimentele toetsvormen. Deze zal als basis worden gebruikt voor het inrichten van instructie en toets waarbij rekening wordt gehouden dat er diverse beheersingsniveaus mogelijk zijn binnen het construct. In hoofdstuk 4 zal de constructie van de toets resulteren in drie toetstaken die gerelateerd zijn aan het te leren construct. Meer specifiek toetsen deze de drie deelconstructen van het construct. Er wordt rekening mee gehouden dat gelijkend de onderzoeksresultaten van Martinez (1998) de toetstaken kunnen verschillen in moeilijkheidsgraad.

HOOFDSTUK 3

Evaluatie en bijstelling van de instructie

In dit hoofdstuk wordt het oplossen van het toetsprobleem voorbereid door de bestaande instructie zo aan te passen dat het mogelijk wordt om op een andere manier te gaan toetsen dan op basis van door de studenten geproduceerde tekeningen. Als voorloper op de uiteindelijk beoogde oplossing in de vorm van een CBT-toets, wordt daarvoor als eerste aanzet een papieren toets ontwikkeld. De instructie wordt aangepast om met deze nieuw ontwikkelde toets een samenhangend onderwijsarrangement te vormen.

3.1 TOELICHTING OP EVALUATIE BESTAANDE INSTRUCTIE

Na het vooronderzoek, gepresenteerd in de voorgaande hoofdstukken, kan het onderzoeksproces volgens de principes van ontwerpgericht onderzoek worden vervolgd met de volgende fase: de prototypische fase. Deze fase kenmerkt zich als een ontwerpfase waarin twee onderzoekslijnen starten die gezamenlijk bijdragen aan het beantwoorden van de onderzoeksvraag.

De eerste onderzoekslijn start met een analyse van de instructie 'Anatomie rat' om te komen tot verbeterpunten in de huidige instructie. De tweede onderzoekslijn start met de ontwikkeling van een eerste toetsversie volgens een stappenplan voor toetsconstructie (Oosterveld en Vorst, 1998) ontwikkeld in samenspraak met de docenten. Zowel de verbeterpunten voor de instructie, de ontwikkeling van de toets als de literatuurbevindingen over het belang en de rol van taakgerichte opdrachten en feedback leiden tot een aangepaste instructie die aansluit op de toetsvorm. Ten behoeve van de eerste onderzoekslijn worden beide instructies (huidige en aangepaste) in een experiment met elkaar vergeleken en met de, uit de tweede onderzoekslijn, ontwikkelde papieren toets getest. Deze evaluatie levert informatie op over de geschiktheid van de instructies (eerste onderzoekslijn) en de aansluiting op de te ontwikkelen toetsvorm met CBT-items (tweede onderzoekslijn).

3.2 1^E ANALYSE ANATOMIE RAT

Het dissectiepracticum 'Anatomie rat' is geanalyseerd met gebruikmaking van het 'Countenance model' van Stake (1967) dat dienst deed als raamwerk. Het countenance model is een curriculum-evaluatiemodel dat gebruikt wordt om een curriculum in kaart te brengen. Toepassing van dit model leidt tot een overzicht op basis waarvan ontwerpkeuzes kunnen worden gemaakt. Het model onderscheidt de evaluatie in drie fasen, te weten een antecedentfase, een transactiefase en een resultatenfase. De antecedentfase beschrijft de condities voorafgaand aan de instructie. De transactiefase beschrijft de processen gedurende de instructie. De resultatenfase beschrijft de effecten van de instructie. Elk van de drie fasen is onderverdeeld in enkele categorieën. De antecedenten betreffen het docententeam, de studenten, de inhoud van het curriculum, het instructiemateriaal, de leeromgeving en de onderwijsdoelen. De transacties betreffen de communicatiepatronen in de onderwijssetting, de tijdsindeling, het verloop van de gebeurtenissen en het sociale klimaat waarin die gebeurtenissen plaats hebben. De resultaten betreffen de opgedane kennis en vaardigheden en de studiehouding van de studenten. In het raamwerk worden de fasen inclusief categorieën weergegeven in de rijen. Vervolgens zijn er vier kolommen in het raamwerk die twee doelen dienen: beschrijven (de eerste twee kolommen) en beoordelen (de laatste twee kolommen). De eerste kolom wordt gebruikt voor het beschrijven van de gewenste situatie. De tweede kolom is voor het beschrijven van wat feitelijk plaatsvindt. De derde kolom specificeert waaraan de feitelijke situatie zou moeten voldoen. De daarvoor gebruikte normen kunnen afkomstig zijn uit de onderwijssetting zelf, bijvoorbeeld ten aanzien van de onderwijsvorm of de onderwijsdoelen, maar ook door de evaluator van buiten worden ingebracht. In de vierde en laatste kolom volgt de beoordeling die volgt door per fase te kijken naar de overeenstemming tussen de beoogde situatie, de waargenomen situatie en de normen. Daarbij moet ook tussen de fasen worden gecontroleerd of deze op elkaar zijn afgestemd.

Voor de analyse van de instructie zijn de fasen en de categorieën gehandhaafd. De kolommen zijn teruggebracht naar drie door het verwijderen van de kolom voor de norm die specificeert waaraan de feitelijke situatie moet voldoen. De reden daarvoor is dat de inrichting van de instructie stoelt op een jarenlange traditie die ook min of meer tot norm verworden is. Deze tekentraditie heeft er toe geleid dat met onderzoek gestaafe normen daarbij onvoldoende ontwikkeld zijn (Jones,

Reed & Weyers, 1994, 1998; Gaupp, 1904). Voor de normering is om die redenen in het hier beschreven onderzoek uitgegaan van de beoogde situatie als norm. Het toevoegen van een normeringskolom zoals bedoeld in het Stakes Countenance model is om die reden overbodig.

Voor de analyse zijn een zestal onderzoeksactiviteiten uitgevoerd. Ze worden hieronder kort beschreven.

Wekelijkse werkbesprekingen met een vakdocent

Deze vonden plaats gedurende zeven maanden waarbij een tweede vakdocent driewekelijks aansloot. De werkbesprekingen met de docent(en) waren gericht op het krijgen van inzicht in de instructie (Anatomie van de rat) en de toetsproblematiek. Centraal daarin stonden vragen als: wat moeten de studenten leren (onderwijsdoelen), wat leren ze daadwerkelijk, hoe wordt dat getoetst, toetst men daadwerkelijk wat men beoogt en welke problemen doen zich daarbij voor.

Uitvoeren van een dissectiepracticum door de onderzoeker

Het uitvoeren van een dissectiepracticum door de onderzoeker had een informatief doel om meer grip te krijgen op het onderzoeksonderwerp en de vraagstukken. Het leidde onder meer tot meer informatieve discussies in de werkbesprekingen met de docent(en).

Observatie van een regulier dissectiepracticum met studenten

Observeren van de instructie omvatte onder meer het volgen van de tijdsplanning en het procesverloop: hielden de assistenten zich aan de afgesproken kaders, hoe verliep de communicatie, hoe werd het instructiemateriaal gehanteerd, wat is het verloop van gebeurtenissen en welk sociaal klimaat heerste er.

Een drietal werkbijeenkomsten met studentassistenten

De werkbijeenkomsten met de studentassistenten waren gericht op het observeren van de docent die de studentassistenten voorbereidde op hun taken en de wijze waarop de studentassistenten daarmee omgingen. Centraal stond onder meer of de studentassistenten de materie beheersten, hoeveel eigen inbreng ze hadden in de didactiek of dat ze moesten werken volgens een vaste procedure, of ze geïnstrueerd werden in de correctiemethodiek van de tekeningen,

Observatie van de eindbeoordelingsbespreking tussen de vakdocent en de studentassistenten
Tijdens het eindbeoordelingsgesprek is vooral gelet op de cijfers die zijn toegekend met de bijbehorende argumentatie en de wijze waarop deze eventueel werden bijgesteld.

De data die deze onderzoeksactiviteiten hebben opgeleverd zijn vastgelegd in logboeken. Deze zijn desgewenst opvraagbaar bij de auteur.
Het resultaat van deze analyse staat weergegeven in tabel 3.1.

Tabel 3.1 *Analyse dissectiepracticum 'Anatomie rat' n.a.v. Stakes Countenance model*

ANTECEDENTEN	'Anatomie rat'		
	Gewenste situatie	Waargenomen situatie	Beoordeling
	1 Docententeam ▪ 1 vakdocent ▪ Afhankelijk van groeps-grootte 6 à 8 assistenten ▪ Een ervaren docententeam dat eenduidig instrueert en beoordeelt	Docententeam Het team bestond uit één vakdocent en diverse student-assistenten met wisselende onderwijservaring. Vooraf aan de instructie werd als training gezamenlijk het practicum verricht als simulatie van het onderwijs. De instructie verliep in grote lijnen gelijk voor een ieder maar ten aanzien van details werd collegiaal regelmatig om advies gevraagd of meegekeken.	Docententeam Door de ondersteuning van studenten met wisselende onderwijservaring wordt de gewenste situatie van een ervaren docententeam in grote lijnen gehaald. Ten aanzien van de details is meer training vooraf wenselijk
	2 Student ▪ Twee typen studenten ▪ Eerstejaars biologen (ca 70) en tweedejaars farmaceuten (ca 130) ▪ Geen dissectie-ervaring met zoogdieren	Student De groep zoals geobserveerd bestond uit farmaciestudenten zonder enige dissectie-ervaring	Student Voldoet aan de norm.
	3 Curriculum inhoud De rat wordt bestudeerd in 10 instructiethema's waarvan 9 dissectielagen	Curriculum inhoud De inhoud van het curriculum is behandeld zoals voorgenomen	Curriculuminhoud Voldoet aan de norm
	4 Instructiemateriaal/bronnen ▪ Een rat (kort vooraf gedood) ▪ Instructiesyllabus ▪ Dissectiegereedschap (schaar, pincet)	Instructiemateriaal/ bronnen ▪ Een rat (kort vooraf gedood) ▪ Instructiesyllabus ▪ Dissectiegereedschap (schaar, pincet) • Sommige assistenten hadden ook hun eigen eerder gemaakte tekeningen meegenomen of andere	Doordat assistenten als begeleiders eigen materiaal meenemen ontstaan er verschillen tussen studenten in de informatie die ze ontvangen. Voorts stellen de assistenten met extra tekeningen impliciet een eigen

Tabel 3.1 *Analyse dissectiepracticum 'Anatomie rat' n.a.v. Stakes Countenance model*

'Anatomie rat'			
	<i>Gewenste situatie</i>	<i>Waargenomen situatie</i>	<i>Beoordeling</i>
		tekenvoorbeelden zoals de 'beste' en de 'slechtste'. De aanpak in het geven van de instructie tussen assistenten verschilt weleens.	norm voor ten aanzien van de tekenkwaliteit die niet over alle studenten consequent wordt doorgevoerd..
	5 <i>Context / omgeving</i> - Dissectie/laboratoriumlokalen - Levendige omgeving waarin veel overleg kan plaatsvinden.	<i>Context / omgeving</i> Dissectie vond plaats in twee laboratoriumlokalen. De omgeving was inderdaad levendig maar ook bijna rumoerig vanwege continu overleg. Studenten die eerder klaar waren in vergelijking tot anderen in de les gingen sociale gesprekken aan wat weleens als storend werd ervaren.	Tijdens de instructie lukt het een levendige omgeving te creëren met als bijkomstig nadeel dat de onderlinge communicatie ook wordt gebruikt voor andere gespreks-onderwerpen dan het studie-object. Het moet mogelijk zijn overleg te hebben over het studieobject zonder te verzanden in sociale gesprekken.
	6 <i>Onderwijsdoelen</i> - Herkennen van anatomische structuren van de rat aan vorm, kleur en textuur - Benoemen van deze anatomische structuren met de Latijnse naam - Aangeven van de ruimtelijke relaties tussen de structuren rekening houdend met relatieve grootte van elke structuur in relatie tot andere	<i>Onderwijsdoelen</i> - De onderwijsdoelen zoals beoogd worden ook geïnstrueerd. - Tijdens de uitleg over relaties tussen objecten werden regelmatig uitstappen gedaan naar de functies van objecten en hun onderlinge functionele relaties, bijvoorbeeld over het mechanisme van het maag-darmkanaal of het mechanisme van de geslachtsorganen - Naast de Latijnse benaming werd veel gebruik gemaakt van de Nederlandse en/of de Engelse benaming	<i>Onderwijsdoelen</i> De doelstellingen worden geïnstrueerd maar besproken dient te worden met de vakdocenten of deze niet dienen te worden verruimd door het betrekken van functionele relaties. Het herkennen van objecten en hun onderlinge ruimtelijke relaties lijkt samen te gaan met het behandelen van de bijhorende functies van de objecten.
TRANSACTIES	7 <i>Communicatielijnen</i> - Tussen vakdocent en assistenten - Assistenten onderling - Docent/assistent en studenten - Studentenkoppels onderling - Onderwerp: dissectie rat	<i>Communicatielijnen</i> De student worden per twee aan elkaar gekoppeld als werkkoppels. Deze kunnen gezamenlijk vakinhoudelijk overleg hebben en feedback leveren op elkaars prestaties. Studenten maken gebruik van de leeromgeving waarin over-	<i>Communicatielijnen</i> In een aangepaste instructie dienen de communicatielijnen te worden gehandhaafd ten behoeve van vakinhoudelijk overleg maar kan worden geoptimaliseerd door mogelijkheden tot sociale communicatie te beperken.

Tabel 3.1 *Analyse dissectiepracticum 'Anatomie rat' n.a.v. Stakes Countenance model*

'Anatomie rat'			
	<i>Gewenste situatie</i>	<i>Waargenomen situatie</i>	<i>Beoordeling</i>
		leg plaatsvindt maar ze bespreken zowel vakinhoudelijke als sociale onderwerpen. Voorts vindt er ook veel communicatie plaats tussen studenten die onderling niet gekoppeld zijn. Deze gaat niet altijd over het dissectiepracticum	
8	<i>Tijdslijn</i> ▪ Duur practicum is twee dagen ▪ Dag 1: Zes instructiethema's waarvan vijf dissectielagen ▪ Dag 2: Vier instructiethema's, met vier dissectielagen ▪ Voor elk thema een variërende duur van 45 minuten tot anderhalf uur ▪ Beoogde planning voor eindbeoordeling is drie weken	<i>Tijdslijn</i> ▪ De duur van het practicum is inderdaad twee dagen met de beoogde verdelingen van thema's maar niet binnen de gestelde lesuren op een dag. De meeste studenten hadden meer tijd nodig en waren elke dag later klaar dan gepland. Dit gaf aan het eind van elke dag de nodige stress voor een aantal studenten. Verschillenden hadden de tekeningen van de eerste dag niet af en moesten die op de tweededag alsnog afmaken ▪ Beoogde planning voor afronding eindbeoordeling is drie weken. Deze planning werd niet gehaald. De studenten kregen hun tekeningen terug na vier weken in plaats van binnen drie weken.	<i>Tijdslijn</i> ▪ In de instructie moet ruimte worden gecreëerd om de tijdsdruk te verminderen. ▪ De studenten worden begeleid tijdens de instructie met mondeling feedback over de dissectie en de wijze van tekenen. Schriftelijk feedback op de geleverde prestaties vindt een aantal weken later plaats. De tijdspanne tussen leer-moment en feedback op de geleverde prestatie is beduidend groter dan wat er beoogd is waarbij de feedback pas komt als het onderwijsblok is afgelopen. ▪ De feedbacktijd van drie weken is onderwijskundig gezien ook te lang (zie feedback, punt 14)
9	<i>Verloop gebeurtenissen</i> ▪ Introductiepraatje door vakdocent aan de start van het practicum. Uitleg over de instructie, materiaal en omgang met dood dier. Daarna verdeling van studenten in koppels en toewijzing student-assistent. ▪ Per thema: de student-assistent introduceert het thema en doet de dissectie	<i>Verloop gebeurtenissen</i> ▪ Introductiepraatje door vakdocent aan de start van het practicum. Uitleg over de instructie, materiaal en omgang met een dood dier. Daarna verdeling van studenten in koppels en toewijzing student-assistent. ▪ Per thema: de student-assistent introduceert het thema en demonstreert de	<i>Verloop gebeurtenissen</i> ▪ De gebeurtenissen tijdens de instructie vinden plaats conform de doelstelling. Er lijkt een aandacht-verschuiving plaats te vinden van 'dissectie en leren door observeren' naar 'de productie van een tekening voor een goede beoordeling'. Dit leidt tot de vraag of een beoordeelde

Tabel 3.1 *Analyse dissectiepracticum 'Anatomie rat' n.a.v. Stakes Countenance model*

'Anatomie rat'			
	<i>Gewenste situatie</i>	<i>Waargenomen situatie</i>	<i>Beoordeling</i>
	van dat onderdeel voor waarbij studenten vragen kunnen stellen. Vervolgens gaan de studenten zelf de dissectie van de betreffende rattenlaag verrichten. Tot slot wordt de dissectielaag getekend. Meer specifiek zal de student elke dissectielaag van de rat met een collega-student ontleden. Vervolgens wordt de blootgelegde laag geobserveerd en vastgelegd in een tekening. De getekende anatomische structuren worden benoemd met de Latijnse benaming. Deze tekening dient daarmee een weergave/vastlegging te zijn van de drie eerder genoemde onderwijsdoelen. ▪ Na afronding van het tweedaags practicum leveren de studenten hun tekeningen in bij de bijbehorende begeleider ter beoordeling. ▪ Na enkele weken ontvangen de studenten de gecorrigeerde tekeningen terug bij wijze van feedback voorzien van een beoordeling.	dissectie waarbij studenten vragen kunnen stellen. Vervolgens gaan de studenten zelf de dissectie verrichten. Tot slot wordt de dissectielaag getekend inclusief het erbij schrijven van de Latijnse namen van de afzonderlijke structuren. ▪ Studenten vergelijken hun tekeningen onderling continue en gedurende de behandeling van de thema's zijn de studenten vooral bezig met de productie van de tekening die voldoet aan de eisen voor een goede beoordeling. ▪ De studenten leveren na het practicum elk de eigen set tekeningen in ter beoordeling en krijgen deze gecorrigeerd voorzien van goed/fout feedback terug na ca vier weken.	tekening het bewijs is dat iemand de leerdoelen heeft behaald (het te leren construct heeft geleerd) of dat de tekening het bewijs is dat iemand voldoende kan tekenen conform de vereisten of een combinatie van beide mogelijkheden. ▪ De tijdsperiode tussen instructie en beoordeling liep vertraging op, de feitelijke beoordelingstijd was ca 135% van de beoogde beoordelingstijd.
	<i>10 Sociaal klimaat</i> Ontspannen sfeer waarin volgens schema wordt gewerkt	<i>Sociaal klimaat</i> ▪ Tijdens het introductiepraatje en de eerste confrontatie met een dode rat zijn sommige studenten emotioneel. Daarvan heeft een enkeling een rat als huisdier. Deze emotionaliteit verdwijnt na behandeling van de eerste thema's. ▪ De sfeer waarin volgens schema wordt gewerkt verloopt redelijk ontspan-	<i>Sociaal klimaat</i> De emotionaliteit die kan ontstaan is een gegeven ervan uitgaand dat dit practicum altijd wordt gegeven met een dissectie. Het negatieve effect van de tijdsdruk dient te worden tegengegaan.

Tabel 3.1 *Analyse dissectiepracticum 'Anatomie rat' n.a.v. Stakes Countenance model*

'Anatomie rat'			
	Gewenste situatie	Waargenomen situatie	Beoordeling
		nen maar om het allemaal af te krijgen ontstaat een tijdsdruk. Aan het eind van de twee dagen zijn de studenten en assistenten vermoeid en is er een tijdsdruk ontstaan om de tekeningen af te krijgen	
RESULTATEN	11 <i>Prestatie student/ beoordeling</i> - De prestatie van de student wordt beoordeeld op basis van de geproduceerde tekeningen. Per student een set van 10. - De beoordeling vindt plaats op netheid, volledigheid van onderdelen, ruimtelijke ligging en verhoudingen tussen onderdelen, correcte Latijnse benamingen van onderdelen. - Beoogd wordt een eenduidige beoordeling door alle beoordelaars afzonderlijk en gezamenlijk	<i>Prestatie student/ beoordeling</i> - De student wordt beoordeeld op de geproduceerde tekeningen waarbij per tekening wordt gelet op netheid, volledigheid van onderdelen, ruimtelijke ligging en verhoudingen tussen onderdelen, correcte Latijnse benamingen van onderdelen. - Elke beoordelaar corrigeert van de eigen studenten de tekeningen. Aan de start van de beoordeling blijken beoordelaars strenger dan aan het eind. Dat wordt gecorrigeerd door na de laatste beoordeling de eerste vijf beoordelingen nog eens te beoordelen - Ter afsluiting van het beoordelingstraject komen alle beoordelaars bij elkaar (vakdocent en student-assistenten) die ervaringen uitwisselen met de beoordeling, elkaars beoordelingen vergelijken op gemiddelde groepsscores, en bij afwijkingen van één of meer punten in gemiddelde groepsscores de cijfers bijstellen naar het algemene gemiddelde waarbij de ervaring van de vakdocent en de gemiddelde beoor-	<i>Prestatie student/ beoordeling</i> De objectiviteit van de toetsingsmethode is niet gegarandeerd. Dat wordt veroorzaakt door de onderlinge verschillen in ervaring tussen de vakdocent en de studentassistenten, de beoordelingscriteria die niet eenduidig zijn vastgelegd en de beoordelaarsamenstelling die verschilt per practicum. Het gevolg is dat bij grote beoordelingsverschillen tussen beoordelaars deze verschillen moeten worden bijgesteld. De methodiek om deze problematiek op te lossen, verhelpt de problematiek niet.

Tabel 3.1 *Analyse dissectiepracticum 'Anatomie rat' n.a.v. Stakes Countenance model*

'Anatomie rat'			
	<i>Gewenste situatie</i>	<i>Waargenomen situatie</i>	<i>Beoordeling</i>
		delingen die hij heeft gegeven aan de tekeningen van zijn groep leerlingen leidend is. Bijvoorbeeld het gemiddelde van de meeste beoordelaars ligt voor de hele groep op 6½, dan zal een beoordelaar met groeps-gemiddelde 7½ in samen-spraak met de vakdocent de tekeningen nogmaals door-nemen en de cijfers naar gelang het oordeel van de vakdocent bijstellen. de vakdocent gebruikt de gemid-delde beoordelingen die hij heeft gegeven aan zijn eigen studenten als referentie.	
	12 Houding Van de studenten wordt een onderzoekende houding verwacht waarin ze door eigen fascinatie voor de anatomische bouw van een zoogdier deze geheel observeren en analyseren	Houding De meeste studenten ontwik-kelen of hebben tijdens het practicum een fascinatie voor het dier als studieobject en vertonen een onderzoekende houding naar de anatomische bouw van een zoogdier waarbij ze vragen stellen en over het onderwerp sparren met elkaar.	Houding De waargenomen situatie is grotendeels gelijk aan de gewenste situatie.
	13 Vaardigheden ▪ Dissectievaardigheden ▪ Omgang met een dood dier	Vaardigheden ▪ De studenten doen dissectievaardigheden op met de schaar en pincet. ▪ In hun omgang met het dode dier verschuift het gemoed t.a.v. het dier van emotioneel naar steeds afstandelijker (rat als studieobject)	Vaardigheden De doelstelling is behaald
	14 Feedback ▪ De studenten krijgen tijdens het practicum feedback van de practicumdocenten naar aanleiding van hun eigen vragen die ze stellen aan de docent.	Feedback ▪ Naast het stellen van vragen aan docenten discussiëren de studentenkoppels onderling veel wat ook als feed-back kan worden gezien. ▪ Bij teruggave van de	Feedback De tijdspanse tussen leer-moment en feedback op de prestatie is erg ruim ten nadele van de effectiviteit van het leerproces. De beoogde situatie is geen

Tabel 3.1 *Analyse dissectiepracticum 'Anatomie rat' n.a.v. Stakes Countenance model*

'Anatomie rat'			
	<i>Gewenste situatie</i>	<i>Waargenomen situatie</i>	<i>Beoordeling</i>
	▪ Bij teruggave van de gecorrigeerde tekeningen krijgen studenten feedback op basis van de beoordeling en op basis van de geïnventariseerde fouten.	tekeningen is de tijdspanne tussen het leermoment en de feedback ca. vier weken.	goede situatie want ten behoeve van de effectiviteit dient feedback zo snel mogelijk te worden gegeven (Wang, Haertel & Walberg, 1990, 1993). Uit praktische overwegingen is gekozen voor een correctieperiode van drie weken maar deze is te lang.

Uit de kolom 'beoordeling' van tabel 3.1 blijkt dat er op veel punten verbeteringen mogelijk lijken. Samengevat gaat het om het volgende.

Het *docententeam* (pt 1) blijkt heterogeen samengesteld waarbij de ervaring in het onderwijzen en beoordelen onderling verschillend is. Sommigen houden zich aan het vastgestelde pakket instructiemateriaal en anderen nemen extra studiemateriaal mee om studenten bewust te maken van de kwaliteit van de tekeningen die van ze wordt verwacht. In de beoordeling zijn de verschillen in ervaring ook merkbaar. De gemiddelde beoordeling van de tekeningen verschilt tussen de studentassistenten die als docent participeren in het docententeam. Na vergelijking van de gemiddelde beoordelingen per tekening, worden beoordelingen van de docenten die gemiddeld het meest afwijken bijgesteld. Richtinggevend daarin is het oordeel van de vakdocent die zich weer baseert op zijn ervaringen en de gemiddelde cijfers die hij heeft toegekend aan zijn studenten voor elke tekening. Is de beoordeling van de studentassistenten buiten een spreiding van +1 of -1 van het gemiddelde van de vakdocent dan wordt deze gecorrigeerd. Dit introduceert het element van willekeur dat illustreert dat de interbeoordelaars-betrouwbaarheid niet optimaal is.

Bij de *communicatielijnen* (pt 7) en het *sociaal klimaat* (pt 10) blijkt er veel onderling overleg te zijn tussen studenten, ook tussen studenten die niet zijn samengevoegd. De gesprekken zijn vooral vakinhoudelijk waarin men met elkaar spart, maar er is ook veel sociale communicatie. Sociale communicatie draagt bij aan het creëren van een ontspannen sfeer in een emotioneel beladen practicum. Dat neemt niet weg dat voor het optimaliseren van de instructie moet worden bekeken op welke wijze effectief kan worden ingespeeld op het vakinhoudelijk discussiëren over de stof en het minimaliseren van niet op het leerdoel gerichte communicatie.

De *onderwijsdoelen* (pt 6) worden zoals gepland behaald waarbij zonder dat het een onderwijsdoel is, de studenten veel informatie opdoen over de functionaliteit van de anatomische structuren. Ze worden daar ook niet op getoetst terwijl ze daar wel de kennis van meekrijgen. Dit is besproken met de vakdocenten en deze gaven aan de onderwijsdoelen en toetsing bij dit vak niet te willen uitbreiden.

De *tijdslijn* (pt 8) geeft weer dat het practicum wel wordt doorlopen in de vooraf geplande dagen maar onvoldoende binnen het bepaalde tijdsbestek op een dag. Daar zat per dag een vertraging in variërend van een half uur tot anderhalf uur. De correctieperiode voor de tekeningen is gepland op drie weken terwijl de tekeningen na vier weken werden geretourneerd. De totale beoordelingstijd is daarmee 135% van de gewenste tijd.

De *prestatie van de student* (pt 11) behoort een weergave te zijn van het te leren construct. Door de tijdsdruk die er ontstaat en doordat het belang van een goede tekening steeds wordt aangegeven, verschuift de focus van de instructie van 'dissectie en leren door observeren' naar 'de productie van een tekening ten behoeve van een goede beoordeling'. Hierdoor ontstaat de vraag of een positief beoordeelde tekening daadwerkelijk een weerspiegeling is van het geleerde construct of dat het weergeeft dat de student goed is in het tekenen of een combinatie van beide. De geschiktheid van de tekening als toetsinstrument staat daarmee ter discussie.

De *feedback* (pt 14) die studenten krijgen gedurende het leerproces vindt voornamelijk individueel plaats door de vragen die ze stellen aan de docent, het onderling over de stof van gedachten wisselen in de studentkoppels en een beoordeling van de door hen geproduceerde tekeningen. Dat laatste is voor verbetering vatbaar omdat ook zonder tijdsoverschrijding van de nu als gewenste termijn geformuleerde drie weken de tijdspanne te lang is.

Bovenstaande elementen geven de aspecten aan die voor verbetering vatbaar zijn of aandacht vereisen. Een aantal daarvan bevestigt de probleemstelling uit hoofdstuk 1. Het gaat daarbij om a) de ontwikkeling van een objectieve toets die b) aansluit op het leerproces en c) studenten in een relatief kort tijdsbestek van feedback voorziet. Nieuwe aspecten zijn d) de communicatie die kan worden geoptimaliseerd, e) de tijdsdruk die een rol speelt, f) de instructie van functionaliteit die nu niet wordt meegenomen als leerdoel en dus ook niet

getoetst wordt en g) de vraag of de geproduceerde tekening een adequaat toetsingsinstrument is. Deze aspecten worden verder in dit hoofdstuk bij de aanpassing van de instructie meegenomen.

3.3 DE TOETS

De toets is ontwikkeld in samenspraak met twee vakdocenten van het practicum 'Anatomie van de rat' met als uitgangspunt de bestaande leerdoelen. Deze leerdoelen zijn, zoals in vorige hoofdstukken aangegeven: (a) het herkennen van anatomische structuren van de rat aan vorm, kleur en textuur, (b) het kunnen benoemen van deze anatomische structuren met de Latijnse naam en (c) het kunnen aangeven van de ruimtelijke relaties tussen deze structuren waarbij rekening wordt gehouden met de relatieve grootte van elke structuur gerelateerd aan andere structuren. De in relatie tot deze doelen opgedane kennis vormt het construct dat een toetskandidaat beheerst. De toets die op basis van deze leerdoelen wordt geconstrueerd, beoogt te meten of het verworven construct ook het gewenste construct is.

Het feit dat, hoewel het geen leerdoel is, de studenten tijdens het practicum veel informatie opdoen over de functies van de diverse anatomische structuren en hun onderlinge functionele relaties, vormde in overleg met de vakdocenten voorsnog geen aanleiding om op dit punt een expliciet leerdoel toe te voegen. De redenering is dat als iemand de ruimtelijke relaties tussen structuren kan aangeven, deze impliciet kennis heeft van de functionele relaties en dus ook van de functionaliteit van de structuren.

Vanuit de toetsconstructie ontwikkelden Oosterveld en Vorst (1998) een stappenplan voor de constructie van vragenlijsten en vertaalden deze naar een zestal methoden waaronder de constructmethode. Deze constructmethode is gericht op het samenstellen van een vragenlijst die een goede operationalisatie is van een onderliggend theoretisch begrip, door hen ook benoemd als "construct". Het stappenplan in de constructmethode is toegepast voor de ontwikkeling van de toets, waar nodig aangepast aan de huidige onderzoekssituatie en omvat de volgende stappen (1) het omschrijven van het te toetsen kennisdomein, (2) het specificeren van item- en testkenmerken, (3) het genereren van een verzameling items evenals het samenstellen van de toets, (4) het relateren van de toets aan de

experimentele vraag en (5) het beschrijven van de psychometrische analyses voor de evaluatie van de toets alsook de toetsing van de experimentele vraag.

In gesprekken met de vakdocenten zijn het construct en de leerdoelen geformuleerd en is het kennisdomein omschreven. Er zijn voor het kennisdomein in totaal 10 instructiethema's, te weten (a) habitus, (b) ventrale spieren, (c) situs viscerum buikholte, (d) uitgelegde darmtractus, (e) vrouwelijke geslachtsorganen, (f) mannelijke geslachtsorganen, (g) kop en hals ventraal, (h) situs viscerum borstholte, (i) grote bloedvaten en organen van de hals en (j) nieren, bijnieren en bloedvaten. De vakdocenten hebben aangegeven belang te hechten aan het toetsen van de mate van herkenning van anatomische structuren. In de literatuur over visuele toetsen (Dwyer, 1978) werd in het PSE-programma gebruik gemaakt van meerkeuzevragen voor herkenning. Ook in het toetsprogramma Ratview (De Graaff en Verhagen, 1997), dat gaat over het instructiethema "mannelijke geslachtsorganen van de rat", wordt getoetst op herkenning waarbij de student uit een aantal anatomische structuren de gevraagde structuren kan kiezen. Met het oog op snelle feedback en een hanteerbare scoringsmethode voor een papieren toets is gekozen voor meerkeuzevragen als itemformat. Daarbij is rekening gehouden met het buitengewone menselijk vermogen om plaatjes te herkennen die eerder zijn gezien (Levie, 1987; Goldstein, 1975) en tot plafondeffecten kunnen leiden.

Voor de hier ontworpen toets is gekozen om de vragen in de stam voornamelijk verbaal op te stellen met soms ondersteuning door een plaatje en de antwoordalternatieven in de vorm van plaatjes over anatomische structuren weer te geven met in een aantal gevallen ook verbale antwoordalternatieven

In de keuze van visuele representaties is gekozen om te variëren in representaties zoals weergegeven in tabel 2.1. Dit is vooral ingegeven doordat de literatuur geen duidelijkheid biedt over welke representaties het effectiefst zijn. Het eindresultaat is een toets bestaande uit 41 tekstuele meerkeuzevragen waarbij de tekstuele vragen al of niet ondersteund worden met plaatjes in de stam. De antwoordalternatieven kunnen zowel tekstueel als met plaatjes zijn.

In de ontwikkeling van de toets is een aantal keuzes gemaakt:

1. De thema's in het practicum dienden evenwichtig te zijn verdeeld over de toets. Bij de constructie van de vragen en de samenstelling van de toets is daar rekening mee gehouden.

2. De aard van de vragen dienen de leerdoelen van het practicum te dekken. Door de toets te ontwikkelen met vakdocenten is bewerkstelligd dat de aard van de vragen de opvattingen van de docenten met betrekking tot de doelen van het practicum weerspiegelt.
3. Vragen die gebruik maken van gedetailleerde plaatjes zijn per kandidaat afwisselend gesteld met een kleurenfoto of met een gedetailleerde zwart-wit tekening. Daarnaast zijn er twee toetsversies waarbij de ene kandidaat de vraag met een kleurenfoto krijgt terwijl de andere kandidaat dezelfde vraag krijgt met een gedetailleerde zwart-wit tekening. Deze varianten zijn gebruikt om een nauwkeurige vergelijking te maken tussen de representatievormen.
4. Bij de gebruikte plaatjes is er een mix van realistische afbeeldingen, contourtekeningen alsook analoge en arbitraire representaties.

De bijlagen horende bij hoofdstuk 3, 3.1 en 3.2, bieden een totaal overzicht van de samenstelling van de toets. De toets dient in dit hoofdstuk als basis voor de uitvoer van het experiment waarin de bestaande instructie wordt vergeleken met een aangepaste.

De beoordelingscriteria zijn in samenspraak met de vakdocent opgesteld. Elk goed ingevuld item werd gescoord met 0 (fout) of 1 (goed). Voor de eindbeoordeling worden alle punten opgeteld, gedeeld door het aantal items en vermenigvuldigd met een factor 10. Dit levert een score op tussen 1 en 10. Deze score geldt als eindcijfer.

Met het inzetten van deze lijn is voor de docenten a priori duidelijk dat het overgaan op een nieuwe toetsmethode de consequentie heeft dat de onderwijsdoelstellingen van het practicum op een nieuwe manier worden geoperationaliseerd. De cursusleiding hoopt te bereiken dat beter meetbaar wordt of de studenten via het dissectiepracticum hebben geleerd hoe een zoogdier is opgebouwd. Zoals eerder genoemd sluit correctie van tekeningen niet uit dat er ruis is in de beoordeling of de onderwijsdoelstellingen al of niet zijn behaald door een combinatie van het meten van het leereffect en de beoordeling van de tekening op netheid en visuele aantrekkelijkheid.

3.4 DE INSTRUCTIE

Het huidige practicum “Anatomie van de Rat”

Het practicum bestaat uit drie taken die de studenten na de algemene verbale instructie verrichten, namelijk (a) een dissectietaak, (b) een tekentaak en (c) een benoemingstaak. In 3.1 is dit inzichtelijk gemaakt. Voor de instructie wordt verondersteld dat de tekentaak de studenten dwingt tot nauwkeurige observatie van de anatomische structuren. Voor de toetsing door beoordeling van de voltooide tekeningen wordt beoogd dat deze weergeven in hoeverre de studenten de in de vorige paragraaf geformuleerde leerdoelen (a) en (c) hebben bereikt, het herkennen van de anatomische structuren en de onderlinge relatie(s) tussen structuren. De benoemingstaak heeft tot doel vast te kunnen stellen of de studenten doel (b) hebben bereikt, het kunnen identificeren van de anatomische structuren.

Door de toetsmethode van het beoordelen van een tekening te vervangen door een andere toetsmethode zoals de papierentoets sluit de instructie niet aan op de toets. Hoewel het belang van de tekentaak door docenten wordt aangegeven, worden naar aanleiding van de analyse in tabel 3.1 ook een aantal kanttekeningen geplaatst zoals het arbeidsintensieve karakter waardoor de uitvoer ervan onder continue tijdsdruk staat. De wijziging in de toetsmethodiek is aanleiding om het verplichte karakter van de tekentaak ter discussie te stellen. Besloten is in een aangepaste instructie de tekentaak weg te laten en te vervangen door een andere. De functie van de tekentaak moet daarbij behouden blijven als het gaat om het nauwkeurig observeren van de verschillende anatomische structuren. De tekentaak is ook nauw verbonden met de benoemingstaak. Het vervangen van de tekentaak in de aangepaste instructie heeft tot gevolg dat ook de benoemingstaak dient te worden gewijzigd. In de volgende paragraaf wordt aan de hand van relevante literatuur ingegaan op de vraag wat beide taken kunnen betekenen voor het visueel leerproces.

Taakgerichte opdrachten in relatie tot visueel leren

In hoofdstuk 2 worden naar aanleiding van het literatuuronderzoek enkele conclusies getrokken:

1. Instructie is noodzakelijk om sturing te geven aan het leerproces ter ondersteuning van de selectie en de organisatie van relevante informatie.
2. Taakgerichte opdrachten zijn hierin een belangrijk instrument. Ze dragen bij aan het richten van de aandacht, maken effectieve selectie mogelijk ten

voordele van de beperkte verwerkingscapaciteit en dragen bij aan het verkrijgen van correcte mentale beelden van de objecten die samen en in samenhang het te leren construct vormen.

3. De didactische kwaliteit van de tekentraditie is vooral gelegen in het feit dat het te beschouwen is als een (component van een) taakgerichte opdracht.
4. De constructie van tekenen als taakgerichte opdracht is effectief mits (a) de waarnemer een instructie vooraf krijgt over de belangrijke kenmerken en hun relaties en (b) de waarnemer het object meerdere malen tekent.

In deze paragraaf worden taakopdrachten verder verkend met specifieke aandacht voor het tekenen en alternatieve taakgerichte opdrachten met als doel een geschikt alternatief te vinden voor het tekenen.

Er zijn kanttekeningen te plaatsen bij de studies die de effectiviteit van het tekenen aantonen.

Ten eerste zijn er verschillende studies die de effectiviteit van het tekenen aantonen gebaseerd op de vergelijking tussen instructies met tekenopdrachten en instructies met sec observeren (Van Meter, Aleksic, Schwartz & Garner, 2006; Van Meter, 2001; Lansing, 1984, 1981). Daarmee is niet aangetoond dat tekenen als taakgerichte opdracht de enige is die bijdraagt aan een goede sturing van het leerproces of dat deze het effectiefst is in vergelijking tot andere taakgerichte opdrachten. Andere onderzoeken waarin tekenen effectief blijkt, baseert de onderzoeksresultaten op het vergelijken van het leren van tekst versus het leren van tekst met een tekenopdracht (Alesandrini, 1981, Dempsey & Betz, 2001; Stein & Power, 1996; Britton & Wandersee, 1997). Ook deze vergelijkingen tonen niet aan dat tekenen de enige effectieve taakopdracht is.

Ten tweede blijkt bij bestudering van onderzoeken die de effectiviteit van het tekenen bepleitten dat er vaak sprake is van de aanwezigheid van andere taakgerichte opdrachten die onderzoekers classificeren als nevenopdrachten (Dempsey & Betz, 2001; Van Meter & Garner, 2005). Zij beschouwen deze als ondersteunend aan de tekenopdracht en kennen de gevonden effecten enkel toe aan het tekenen. Ook de instructie in het hier gepresenteerde onderzoek is niet beperkt tot het enkel construeren van tekeningen want het kunnen benoemen van de getekende onderdelen kan worden beschouwd als een taakgerichte opdracht. Deze kanttekeningen worden duidelijk aan de hand van de onderzoeken die hieronder worden beschreven.

Dempsey en Betz (2001) onderschrijven het belang van tekenen als strategie omdat het lerenden ondersteunt in het signaleren van details en subtiële eigenschappen van het object. Zij adviseerden het gebruik van een didactisch systeem waarbij studenten met vijf oefeningen leren zelf geproduceerde tekeningen te gebruiken als een strategie om observatievaardigheden op te doen. Interessant is dat elke oefening gebaseerd is op een tekentaak in combinatie met andere taken zoals lezen of benoemen of klassikale discussie. Tekenend wordt door deze auteurs gezien als een op zichzelf staande opdracht ondersteund door nevenopdrachten. De mogelijkheid dat de andere opdrachten van vergelijkbaar niveau zijn als het tekenen of een mogelijk even zo grote invloed hebben op het leerproces is niet meegenomen.

Ook Dean en Kulhavy (1981) hebben onderzoek gedaan naar (a) de effecten op het leerproces van het produceren van een grafische representatie naar aanleiding van een tekst en (b) de bepalende factoren voor het leren van een tekst. De bepalende factoren waren (a) een grafische representatie van de tekst, (b) een mondelinge toelichting over de tekst en (c) een benoemingsopdracht van de objecten in de grafische representatie. In het eerste experiment werd een tekst gebruikt over de plaats en leefsituatie van een Afrikaanse stam. De hypothese was dat als de proefpersonen de opdracht krijgen om op basis van die tekst een landkaart te maken dit de begrijpelijkheid van de tekst vergemakkelijkt en daardoor het leerproces positief beïnvloedt. In dit experiment werden 40 proefpersonen verdeeld in twee groepen. De proefpersonen kregen allemaal een onbekende tekst over de topografie en levenswijze van de Afrikaanse stam. Eén groep kreeg de opdracht om de informatie uit de tekst grafisch weer te geven door het tekenen van de landkaart. Deze kaart werd gezien als een spatiële organizer. De andere groep hoefde niet te tekenen maar alleen de tekst te bestuderen. Na de instructie werden beide groepen getoetst. De toets bestond uit drie delen (a) 35 meerkeuzevragen, (b) 35 constructievragen en (c) een free recall deel van 10 minuten. De constructievragen waren kort-antwoordvragen. Het resultaat van het experiment was dat studenten die een tekening hadden gemaakt gemiddeld hoger scoorden op de test dan de studenten die niet hadden getekend. De onderzoekers concludeerden uit dit experiment dat het maken van de tekening studenten dwingt om de aandacht te richten op belangrijke informatie en om deze informatie te organiseren in een ruimtelijke structuur. Deze onderzoeksresultaten hebben geresulteerd in het tweede experiment. Voor het tweede experiment werden 120 studenten verdeeld in drie groepen van 40 studenten. De eerste groep kreeg een

gestandaardiseerde topografische tekening waarbij studenten alle objecten moesten benoemen aan de hand van een mondelinge toelichting over de ruimtelijke relaties tussen de objecten in de tekening. De tweede groep kreeg dezelfde topografische tekening, waarin nu alle objecten door middel van bijschriften waren benoemd. Deze groep kreeg geen toelichting maar werd wel gevraagd de tekening goed te bestuderen. De derde groep kreeg een tekening maar deze had geen relatie tot het onderwerp. Daarna kregen alle deelnemers de tekst die hoorde bij de toelichting. De deelnemers moesten de tekst bestuderen zonder de tekening. Na bestudering van de tekst werd de test aangeboden. De resultaten van het experiment waren dat de groep die met behulp van een tekening objecten moest benoemen en bovendien een mondelinge toelichting hadden ontvangen gemiddeld hoger scoorde op de test dan de groep die een tekening met bijschriften bestudeerde zonder toelichting. Deze groep scoorde gemiddeld wel hoger dan de groep die een tekening had gekregen die geen relatie had tot de instructie. De onderzoekers concludeerden ook uit deze resultaten dat er sturing nodig is om de aandacht te richten en dat dit mogelijk is met taakgerichte opdrachten. Uit hun eerste experiment blijkt dat het produceren van een tekening op basis van de tekst een positiever effectief heeft ten opzichte van alleen het bestuderen van de tekst. Verder kan de opdracht om te benoemen ook als een taakgerichte opdracht worden beschouwd waarbij is aangetoond dat zonder tekeningen een factor zoals benoemen ook een positief effect heeft op het leerproces. In dit geval was het aanbieden van een tekening als spatiële organizer met als opdracht het benoemen van de objecten en een mondelinge toelichting het effectiefst om de ruimtelijke organisatie van objecten te leren.

In recentere literatuur van Van Meter en Garner (2005) pleitten de auteurs voor onderzoek waarbij het tekenen door de lerende wordt beschouwd als een leerstrategie. Daarbij geven ze aan dat (gedwongen) ondersteuning in combinatie met door de lerende geproduceerde tekeningen effectiever is dan alleen de productie van een tekening. De productie van een tekening door de lerende wordt dus beschouwd als een taakgerichte opdracht en het benoemen als een toegevoegde taakgerichte opdracht. Ook Van Meter en Garner beweren dat het effect van tekenen dusdanig is dat alle andere opdrachten daarnaast enkel van ondersteunende aard zijn. Volgens de auteurs vervullen deze ondersteunende opdrachten een drietal functies, te weten: (1) het stellen van een kader voor het produceren van de tekening, (2) het controleren van de accuraatheid van de geproduceerde tekening en (3) het sturen van de aandacht van de lerende voor

wat betreft de essentiële onderdelen en hun onderlinge relaties. De auteurs geven aan dat er nog onderzoek noodzakelijk is om hun bewering te staven. Het is daarmee nog onbekend wat de bijdrage van elke afzonderlijk taakgerichte opdracht is, of één taak zoals het tekenen het effectiefst is, of juist ondersteunend of dat de combinatie juist het meeste effect heeft.

De beschreven literatuur over taakgerichte opdrachten is gestart naar aanleiding van de bijdrage van het tekenen in de instructie van het hier gepresenteerde onderzoek. De literatuur over het belang van tekenen (Dean & Kulhavy, 1981; Lansing, 1981, 1984; Dempsey & Betz, 2001; Van Meter & Garner, 2005) als taakgerichte opdracht wijst uit dat tekenen in vergelijking tot sec observeren of het bestuderen van een tekst een meerwaarde heeft. Deze positieve bijdrage van het tekenen staat niet ter discussie maar sluit niet uit dat tekenen niet als enige een positief effect heeft op het leerproces. De literatuur biedt zelf ruimte voor alternatieven: Tekenen met daaraan toegevoegde ondersteuning in de vorm van instructie of andere taken zoals benoemen zijn in combinatie effectiever dan tekenen alleen. Voorts toont het werk van Dean en Kulhavy (1981) aan dat naast de productie van een tekening ook het aanbieden van een standaardtekening met een benoemingsopdracht bijdraagt aan het sturen van het leerproces.

Deze laatste bevinding is voor het hier gepresenteerde onderzoek relevant omdat, in relatie tot de te construeren geautomatiseerde toets, de beoordeling van de tekening komt te vervallen en de relevantie van het maken van een tekening ter discussie komt te staan. In geval het aanbieden van een tekening met een benoemingsopdracht even effectief is als het produceren van een tekening met benoemingsopdracht, dan sluit het eerste beter aan op een geautomatiseerde toetsingsmethode.

De hier gepresenteerde literatuur wordt gebruikt voor het ontwikkelen van een alternatieve instructie waarin de tekentaak niet meer voorkomt maar andere taakgerichte opdrachten worden toegepast.

In een alternatieve instructie waarin niet wordt getekend zijn er geen tekeningen die kunnen worden gebruikt voor feedback en toetsing. Het belang van feedback wordt bevestigd door ondermeer Wang, Haertel en Walberg (1990, 1993) die onderzoek hebben verricht naar variabelen die van invloed zijn op het leerproces door het bestuderen van 179 documenten (waaronder meta-analyses) over

instructietheorieën en onderwijskundig onderzoek. Dit resulteerde in 228 variabelen. Deze variabelen werden geschaald op een drie puntenschaal: Eén punt voor een variabele die een zwakke invloed heeft op het leerproces, twee punten voor een variabele die een middelmatige invloed heeft op het leerproces en drie punten voor een variabele die een sterke invloed heeft op het leerproces. Vernoemde variabelen, in relatie tot het hier gepresenteerde onderzoek, die een sterke positieve invloed hebben op het leerproces zijn (a) het gebruik van corrigerende feedback en (b) het stellen van vragen door de docent tijdens het leerproces om de lerenden actief te houden. Binnen de instructietheorie is feedback elke communicatie of procedure die wordt gegeven om de lerende te informeren over de accurateheid van diens response. De meest gangbare procedure is het beantwoorden van een instructievraag door de lerende. Deze krijgt dan onmiddellijk feedback op diens respons. In overeenstemming met het onderzoek van Wang, Haertel en Walberg wordt in verschillende instructietheorieën belang gehecht aan feedback (Gagné, 1985; Reigeluth, 1994; Merrill, 1983). Al sinds Gagne (1972) in de “events of instruction” waarin relevante instructieonderdelen worden weergegeven, is feedback een van de vaste componenten van elke instructie. De lerende krijgt tijdens de instructie feedback over zijn prestaties op oefentaken. Er zijn verschillende vormen van feedback (Mory, 1996, 2004) te weten, motiverende feedback (om lerenden aan te moedigen), belonende feedback (als de leerlingen goed presteren) en informatieve feedback (die informeert over de voortgang in het leerproces). Smith en Ragan (1993) geven aan dat informatieve feedback in een instructie van belang is omdat de lerende dan niet alleen informatie krijgt over zijn respons maar daardoor ook zijn respons kan bijstellen. Zij adviseren, gebaseerd op Gagné (1972), om bij het inrichten van een instructie directe feedback aan te bieden aan de lerende op basis van zijn prestaties op oefentaken zodat deze zijn leergedrag waar nodig kan bijstellen. Ook hierin kan de relatie gelegd worden met de CTML-theorie (Mayer, 2005) van informatieverwerking. Door het principe van actief leren past de lerende cognitieve processen toe op inkomende informatie om zinvolle informatie-eenheden te genereren. Door het aanbieden van directe feedback wordt het actief leren bevorderd en kan de lerende het vormen van zinvolle informatie-eenheden bevorderen en waarnodig bijstellen.

Zo ook recent onderzoek van Smith et al (2009) onder studenten naar het effect van discussiëren in de collegezaal over de lesstof dat uitwees dat onderling discussiëren het begrip over de lesstof bevordert zelfs als de aanwezige kennis onderling onvoldoende is. Zij onderzochten onder 350 studenten hun kennis over

de lesstof. Eerst stemden de studenten individueel op een tentamenvraag. Vervolgens discussieerden ze in kleine groepjes over de stof zonder feedback van de docent en beantwoordden dezelfde vraag nogmaals individueel. Het gemiddeld aantal goede antwoorden steeg van 52 naar 68 procent. Bij een vergelijkbare tweede vraag gaf 77 procent van de studenten het juiste antwoord waarbij van de studenten die de eerste vraag tweemaal fout hadden toch 44 procent de tweede vraag alsnog goed beantwoordden. Hun conclusie was dat studenten de stof beter begrepen na de groepsdiscussie en de kennis pas konden toepassen bij een vergelijkbare nieuwe opgave. Groepsdiscussie kan in dit verband als een volwaardige taakgerichte opdracht worden gezien en ook als vorm van feedback.

Bovenstaande literatuur over taakgerichte opdrachten, het aanbieden van tekeningen, het belang van een benoemingsopdracht en directe feedback bieden voldoende mogelijkheden voor het ontwerpen van een alternatieve instructie, beschreven in de volgende paragraaf, die tegemoet komt aan de aspecten uit tabel 3.1 die kunnen worden geoptimaliseerd en waarin het handmatig produceren van een tekening wordt weggelaten.

Ontwerp van de vernieuwde instructie

Op basis van de analyse van het vak 'Anatomie van de rat' en de argumenten uit de theorie die beschreven zijn in de voorgaande paragrafen kan de vervanging van de tekentaak en de bijbehorende benoemingstaak vorm krijgen. Voor een geautomatiseerde toetsingsmethode waarin het corrigeren van tekeningen irrelevant is, staat het belang van een zelf geproduceerde tekening ter discussie. Voor het ontwerpen van een alternatieve instructie, die aansluit op de nieuwe toetsingsmethode, heeft dit geleid tot het besluit om in plaats van zelf tekenen de studenten in het practicum over de anatomie van de rat van elke anatomische laag een lijntekening ter beschikking te stellen waarin de contouren van de te observeren anatomische structuren zijn getekend. De opdracht is om elk van die structuren te benoemen met een bijschrift. Als vervanging voor de feedbackfunctie, die in de oorspronkelijke instructie werd vervuld door het corrigeren van de tekeningen na het practicum, is gekozen voor een vorm van groepsfeedback tijdens het practicum. Nadat studenten een anatomische laag hebben ontleed en benoemd via bijschriften op de voorgedrukte tekening, volgt een groepsbespreking waarin de benoemingstaak in de groep herhaald wordt. Dit is het moment van terugkoppeling tussen de student en docent en de studenten onderling. De docent heeft een poster van de tekening en geeft elke student een

opdracht waarbij in de tekening een anatomisch object wordt benoemd. Als de student een fout maakt, dan wordt deze verbeterd door de docent of door een andere student. Door deze veranderingen bestaat de aangepaste ‘vernieuwde’ instructie uit drie taakgerichte opdrachten namelijk (a) de dissectietaak, (b) de benoemingstaak en (c) directe feedback in een groepsbespreking.

Figuur 3.1 geeft een overzicht van beide instructievormen.

Figuur 3.1 Overzicht van de bestaande en de aangepaste ‘vernieuwde’ instructie

Met de ontwikkeling van de aangepaste instructie moet worden onderzocht of deze even effectief is als de bestaande. Dit heeft geresulteerd in de volgende vraag:

Is de vernieuwde instructie gelijkwaardig aan de bestaande instructie voor het bereiken van de onderwijsdoelen van het dissectiepracticum anatomie van de rat?

In een experiment worden beide instructies met elkaar vergeleken in het resterend deel van dit hoofdstuk.

Vergelijking van de instructies

Om na te gaan of de aangepaste instructie een goede vervanger is voor de oorspronkelijke instructie, is een experiment uitgevoerd. Er is gewerkt met twee experimentele condities (X_1 en X_2) en één controleconditie (X_0). De experimentele condities verschillen met betrekking tot de taakgerichte opdrachten die er deel van uitmaken. De oorspronkelijke instructie (X_1) kenmerkt zich door dissectie in combinatie met tekenen en benoemen. De aangepaste instructie (X_2) kenmerkt zich door dissectie in combinatie met benoemen en groepsfeedback. Om na te gaan of taakgerichte opdrachten als zodanig een effect hebben op het leerproces, werd in de controleconditie (X_0) alleen de dissectietaak uitgevoerd.

Figuur 3.2 geeft een overzicht van de experimentele condities

Figuur 3.2 Overzicht van de experimentele en controle condities

Als taakgerichte opdrachten bijdragen aan het richten van de aandacht en op deze wijze het leerproces positief beïnvloeden, dan moeten studenten in de condities met opdrachten gemiddeld hoger scoren op de test dan studenten in de kale dissectieconditie. De eerste onderzoekshypothese luidt:

H1: de gemiddelde scores op de toets voor de studenten die de condities X_1 en X_2 hebben gehad, liggen hoger dan de gemiddelde score op de toets voor studenten die de X_0 -conditie hebben gehad. ($X_1 > X_0$; $X_2 > X_0$)

Omdat het gaat om het ontwikkelen van een instructievorm die zich leent voor geautomatiseerde beoordeling van het leerresultaat (en als zodanig geschikt is als testbed voor de experimenten van de hoofdstukken 5 en 6), is van belang dat de alternatieve instructie (X_2) minstens zo goed is als de oorspronkelijke instructie (X_1). De tweede onderzoekshypothese luidt:

H2: de gemiddelde score op de toets voor de studenten met de conditie X_2 is minstens zo hoog als de gemiddelde score op de toets voor de studenten met de conditie X_1 . ($X_2 \geq X_1$)

Hoewel de bestaande instructie volgens de huidige methodiek tekeningen oplevert ter beoordeling, wordt de beoordeling van de tekeningen niet in het onderzoek meegenomen. Redenen daartoe zijn:

- Dat van meet af aan duidelijk is dat deze beoordelingswijze niet zal worden gehandhaafd.
- Dat de relatie netheid tekening en toetsen construct te onzeker is om de beoordeling van de tekeningen te kunnen gebruiken om conclusies te trekken over de papieren toets.

Het onderzoek is uitgevoerd onder biologiestudenten in het eerste jaar van hun studie. Deze studenten hadden geen dissectievaardigheden opgedaan met dieren.

Volgens Campbell en Stanley (1981) komt het de betrouwbaarheid van de resultaten ten goede als op meerdere momenten wordt getoetst. Zij adviseren retentietoetsen een maand later, een half jaar later en een jaar later. Om praktische redenen zijn de meetmomenten in dit experiment: (a) onmiddellijk na het practicum, (b) een week na het practicum en (c) drie weken na het practicum. De voornaamste reden was het zoveel mogelijk beperken van drop-outs omdat het gaat om eerstejaars studenten waarvan een aantal in dat eerste jaar besluit te stoppen met de studie. Om de meeste studenten te bereiken zonder dat deze

afwisten van de test zijn de tijdstippen geselecteerd op momenten waarvan werd verwacht dat de meeste studenten aanwezig zouden zijn op colleges. De studenten zijn niet geïnformeerd over de natoetsen waarmee kon worden voorkomen dat zij de stof nogmaals zouden bestuderen.

Methode en procedure

Randvoorwaarden voor het experiment

Aangezien het experiment werd uitgevoerd in de bestaande onderwijssetting, moeten deze op dusdanige wijze worden uitgevoerd dat voor elke conditie de leerdoelen werden behaald. De deelnemers mochten niet benadeeld worden in het bedoelde leerproces door de instructievorm in de conditie waaronder ze vielen. Een gevolg daarvan was dat het onderwijspersoneel niet kon toestaan dat studenten gedurende het gehele practicum alleen in de dissectie-only-conditie (X_0) zouden werken omdat dat nadelig zou kunnen zijn voor het leerproces van de deelnemers in die conditie. Dit heeft geleid tot een werkwijze waarbij het experiment is gesplitst in twee afzonderlijke designs die betrekking hebben op twee verschillende helften van het practicum. In de praktijk ging dat over twee elkaar opvolgende dagen. De instructie over de anatomie van de rat bestond uit tien thema's van oplopende moeilijkheidsgraad. Op de eerste dag werden de vijf gemakkelijkste thema's behandeld. Voor die dag werd de dissectie-only conditie toegestaan met de verwachting dat de studenten in de dissectie-only conditie weinig nadeel zouden ondervinden door hen bij de gemakkelijke thema's de taakgerichte opdrachten te onthouden. Hierbij hoort onderzoeksdesign 1, zie figuur 3.3, waarin alle drie experimentele condities voorkomen. Op de tweede dag werden de vijf moeilijkste thema's behandeld. De docenten vonden het onverantwoord en ten nadele van de studenten om de dissectie-only conditie te handhaven. De studenten die op de eerste dag de conditie dissectie-only volgden werden at random verdeeld over de andere twee experimentele condities. Hierdoor konden op de tweede dag X_1 en X_2 vergeleken worden met meer proefpersonen per conditie. Het onderzoeksdesign dat bij deze beschrijving hoort staat weergegeven in figuur 3.4.

Onderzoeksdesign

De condities in de eerste instructiehelft resulteerden in het eerste onderzoeksdesign, zie figuur 3.3, waarin twee experimentele condities werden vergeleken namelijk conditie X_1 (dissectie + tekenen + benoemen), conditie X_2 (dissectie + benoemen + groepsfeedback) en de controleconditie X_0 (dissectie only). Het is een posttest-only control group design met gerandomiseerde

toewijzing (R) van proefpersonen aan de condities (Campbell en Stanley, 1981). Er zijn drie toetsen: natoets O1.1 direct na afloop van de instructie, retentie toets O1.2 na één week en retentie O1.3 na drie weken.

R	X ₁	O ₁₁	O ₁₂	O ₁₃
R	X ₂	O ₁₁	O ₁₂	O ₁₃
R	X ₀	O ₁₁	O ₁₂	O ₁₃

Figuur 3.3 Onderzoeksdesign 1: posttest-only control group design voor instructiehelft 1

Voor de condities in de tweede instructiehelft werd de controleconditie X₀ uit de eerste instructiehelft gesplitst in twee gelijke groepen waarna één groep de conditie X₁ onderging en de andere groep de conditie X₂, zie figuur 3.4. Beide groepen hebben hetzelfde voortraject (X₀ en O₁₁) waarvan werd aangenomen dat het voortraject in beide groepen van gelijke invloed zou zijn op de testresultaten. Ook hier is gerandomiseerd (R) en zijn er drie meetmomenten: direct na afloop van het practicum (O₂₁), na één week (O₂₂) en na drie weken (O₂₃).

Voortraject			Design Instructiehelft 2			
R	X ₁	O ₁₁	X ₁	O ₂₁	O ₂₂	O ₂₃
R	X ₂	O ₁₁	X ₂	O ₂₁	O ₂₂	O ₂₃
R	X ₀	O ₁₁	X ₁	O ₂₁	O ₂₂	O ₂₃
R	X ₀	O ₁₁	X ₂	O ₂₁	O ₂₂	O ₂₃

Figuur 3.4 Onderzoeksdesign 2 voor instructiehelft 2

De opzet van de onderzoeksdesigns behoren interpretatie-exclusief (Swanborn, 1993) te zijn, dat betekent dat de onderzoeksresultaten inderdaad het gevolg zijn van de experimentele condities. Toch zijn er factoren die naast de experimentele condities van invloed kunnen zijn op de onderzoeksresultaten (Campbell & Stanley, 1981). Enkele factoren die van invloed zouden kunnen zijn, zijn (a) geschiedenis, (b) bias, (c) uitval en (d) interactie-effect van het testen. De geschiedenis van de proefpersonen is van belang tussen de eerste en tweede test. Door de test eerder aan te bieden in een voortoets kan dat de resultaten van de natoets beïnvloeden. In dit onderzoek is er gekozen voor onderzoeksdesigns met enkel natoetsen zodat het effect van een voortoets niet meegenomen wordt in de natoets. Door de proefpersonen at random toe te wijzen aan de verschillende condities, is een voortoets overbodig omdat ervan wordt uitgegaan dat de gevormde groepen gelijkwaardig zijn. In het tweede onderzoeksdesign is er wel

sprake van een test in het voorttraject, namelijk de test over de eerste instructiehelft, die invloed zou kunnen hebben op de resultaten van de natoets op de tweede instructiehelft. De invloed van de test over de eerste instructiehelft hoeft niet storend te zijn op de testresultaten van de test over de tweede instructiehelft omdat de instructiethema's verschillend zijn. Als de test over de eerste instructiehelft wel van invloed is op de testresultaten van de tweede instructiehelft dan is het te verwachten dat deze gelijk is verdeeld over de twee experimentele condities. Door randomisering van proefpersonen is ook de mogelijkheid van bias geminimaliseerd. Een derde storende factor kan uitval in respondenten zijn. Dit is een factor waarmee rekening moet worden gehouden omdat ervoor is gekozen om studenten niet te vertellen over de retentietoetsen opdat ze zich daar niet op konden voorbereiden. Studenten werden tijdens andere lessen of na een toets "verrast" met een retentietoets. Dit heeft tot gevolg dat er een ongemerkte selectie kan hebben plaatsgevonden in studenten die gemiddeld "meer" interesse hebben in de leerstof dan andere studenten. De vierde factor die de testresultaten kan beïnvloeden is het interactie-effect van testen dat veroorzaakt wordt door het meerdere malen afnemen van de natoets. Het is mogelijk dat studenten de retentietoetsen invullen op basis van herkenning in eerdere testen. Er is getracht dit effect te voorkomen door in de tests de vraagvolgorde te veranderen en de volgorden in antwoordalternatieven te veranderen.

Onderzoeksgroep

De onderzoeksgroep bestond uit 140 eerstejaars biologiestudenten waarvan het aantal studenten dat deelnam aan de retentietoetsen daalde tot 131 tijdens de tweede test en 74 studenten tijdens de derde test. De reden van deze daling is dat de natoetsen in het reguliere onderwijs zijn uitgevoerd waar er geen aanwezigheidsplicht is. Studenten zijn vooraf niet ingelicht over de natoetsen zodat ze zich er niet op konden voorbereiden. Geen van de studenten had voor het experiment een instructie gehad in de dissectie van een zoogdier maar ze hadden wel dissectievaardigheden opgedaan met vissen.

Procedure

Het experiment is uitgevoerd in het eerste curriculumjaar van de opleiding biologie bij de afdeling dierfysiologie. Studenten kregen voor de instructie een introductie met informatie over de dissectie-instrumenten, de practicumprocedure en het veiligheidsreglement. Deze studenten waren vooraf at random toegewezen aan een conditie. De studenten hadden geen voorkennis over de experimentele condities. Voor alle groepen duurde de instructie twee elkaar

opvolgende dagen van elk acht uur. Per dag werd één van de instructie helften doorlopen en de dag werd beëindigd met een test over de die dag behandelde leerstof. Per groep werden de studenten onderverdeeld in groepen van 14 tot 16 studenten met een eigen student-assistent als docent. In totaal waren er acht studenten-assistenten die de instructie gaven. Deze assistenten kregen vooraf aan de instructie een training van de docent over het begeleiden van het practicum en werden geïnformeerd door de onderzoeker over de organisatie en inrichting van het practicum in relatie tot de experimentele condities en de toetsing daarvan.

Data verzamelen

De samengestelde toets bevatte 41 meerkeuze-items en bestond uit twee deeltesten voor elke instructiehelft waarvan 21 items over de eerste instructiehelft (vijf thema's tijdens dag 1) en 20 items over de tweede instructiehelft (vijf thema's tijdens dag 2). Deze is weergegeven in bijlage 3.1 in de bijlagen horende bij hoofdstuk 3. De toets bestaat uit twee toetsversies met als verschil dat als de ene kandidaat een vraag met een kleurenfoto krijgt, de andere kandidaat dezelfde vraag krijgt met een gedetailleerde zwart-wit tekening en vice versa, weergegeven in bijlage 3.2. Tijdens de samenstelling van de test is ervoor gezorgd dat de vragen evenwichtig verdeeld zijn over de thema's van het practicum. Dezelfde test werd op alle drie meetmomenten aangeboden maar per meetmoment verschilden de volgorde van de vragen en de volgorde van de antwoordalternatieven. In de eerste meting werd de test in twee deeltesten afgenomen, dekkend voor de op die dag doorlopen instructiethema's. De studenten maakten, op het tweede en derde meetmoment, alle vragen over het gehele practicum in één test. Bij de data-analyse werden de resultaten op de items over beide practicumhelften (dag 1 en dag 2) apart behandeld en getoetst volgens de onderzoeksdesigns van de figuren 3.3 en 3.4.

Ten behoeve van de beoordeling moesten de studenten hun toetsantwoorden invullen op een optisch leesbaar toetsformulier. Deze ingevulde formulieren werden vervolgens automatisch gescand waardoor de scores werden gedigitaliseerd naar een databestand dat als basis diende voor de verdere verwerking.

Hoewel het experiment primair gericht is op de effectiviteit van de instructie is bij de uitvoering van het experiment ook in de gaten gehouden hoeveel tijd de studenten nodig hadden om de taken uit de experimentele conditie uit te voeren. Als extra variabele is de tijd dus meegenomen en daar wordt in de discussie kort op teruggekomen.

Data analyse

Het significantieniveau voor alle testen was $p = 0,05$. Om de verschillen tussen de experimentele condities te meten was de gemiddelde testscore voor elke conditie berekend en daarop was, afhankelijk van het aantal factoren, een T-toets of een ANOVA-analyse uitgevoerd om na te gaan of eventuele verschillen significant zijn. De retentie werd bepaald door de verandering in de individuele scores van studenten op de natoetsen. Voor de afzonderlijke natoetsen is een Repeated Measures Analyse uitgevoerd. De resultaten zijn gebaseerd op gemiddelde testcores met een schaal van 0 tot 1. De analyses van de retentietoetsen zijn gebaseerd op enkel die studenten die aan alle testen hebben deelgenomen.

3.5 RESULTATEN

Resultaten over onderzoeksdesign 1

In tabel 3.2 is een overzicht gegeven van de gemiddelde scores op de natoets ($O_{1.1}$ in figuur 3.2) met de bijbehorende standaarddeviaties. Er is een significant verschil gevonden tussen de experimentele condities op de test met $F(2,137) = 8,00$ met een Significantieniveau van $p < 0.01$. Dit verschil wordt veroorzaakt door het verschil in gemiddelde testscore tussen de condities X_1 en X_2 met de controle conditie waarbij de gemiddelde testscore van de controlegroep significant kleiner is. Dit is te zien in figuur 3.5 die een schematische weergave is van de gemiddelde testcores op de condities. De ANOVA-analyse wordt weergegeven in tabel 3.3

Tabel 3.2 *Gemiddelde testcores (M) en standaarddeviaties (SD) voor de experimentele condities op de natoets van alle studenten (N=140) die aan het practicum hebben deelgenomen*

Experimentele condities	M	SD
X_1	0,817	0,132
X_2	0,849	0,100
X_0	0,748	0,142

Noot: X_1 : experimentele conditie 1: oorspronkelijke instructie (dissectietaak, tekenen en benoemen); X_2 : experimentele conditie 2: aangepaste instructie (dissectietaak, benoemen en groepsbespreking); X_0 : controle conditie: dissectietaak.

Tabel 3.3 *Anova- analyse van de experimentele condities*

	df	Variantie	F	Overschrijdingskans
Tussen groepen	2	0,136	8,000	0,001
Binnen groepen	137	0,017		

Noot: Significantieniveau ($p \leq 0,05$).

De toetsen zijn na de natoets nog tweemaal aangeboden als retententietoetsen. De resultaten daarvan zijn te zien in tabel 3.4 waarin van de drie tests de gemiddelde testcores zijn weergegeven en de bijbehorende standaarddeviaties. In figuur 3.3 zijn de resultaten in de onderzoeksdesign gesymboliseerd met O1.1, O1.2 en O1.3.

Tabel 3.4 *Gemiddelde testcores (M) en standaarddeviaties(SD) voor de experimentele condities op de natoets en retententietoetsen van studenten (N=74) die alle drie testen hebben gemaakt*

Experimentele condities		M	SD
X1	Natoets	0,809	0,127
	Retententietoets 1	0,717	0,154
	Retententietoets 2	0,684	0,145
X2	Natoets	0,862	0,109
	Retententietoets 1	0,801	0,143
	Retententietoets 2	0,742	0,146
X0	Natoets	0,769	0,155
	Retententietoets 1	0,772	0,149
	Retententietoets 2	0,724	0,128

Noot: X1: experimentele conditie 1.; dissectietaak, tekenen en benoemen; X2: experimentele conditie 2.; dissectietaak, benoemen en groepsbespreking; X0: controle conditie3.; dissectietaak.

Uit de resultaten van de repeated measures analyse in tabel 3.5 blijkt dat er verschillen zijn in gemiddelde scores tussen de tijdstippen waarop de testen zijn afgenomen ($F = 23,749$, $p < 0,01$) en dat er ook sprake is van een interactie-effect ($F = 3,089$, $p < 0,05$) tussen de verschillende tijdstippen van testafname en de experimentele condities. Dit effect is zichtbaar daar voor alle condities geldt dat naarmate de tijd verstrijkt de testresultaten afnemen. Om een beeld te geven van deze verschillen zijn de gemiddelde scores per testtijdstip en conditie weergegeven in figuur 3.5. Te zien is dat de scores op elke toets in de tijd bij elke meting afnemen.

Noot: X1: experimentele conditie 1,; dissectietaak, tekenen en benoemen; X2: experimentele conditie 2; dissectietaak, benoemen en groepsbespreking; X0: controle conditie3; dissectietaak.

Figuur 3.5 Gemiddelde testscores voor de experimentele condities op de natoets en retentietoetsen (N=74)

Tabel 3.5 *Repeated measures analyse voor de experimentele condities op drie tijdstippen*

	Df	Variantie	F	Overschrijdingskans
Testtijdstippen	2	0,159	23,749	0,000
Interactie tijdstippen en condities	4	0,020	3,089	0,018
Conditie	2	0,063	1,352	0,265

Noot: Significantieniveau ($p \leq 0,05$).

In figuur 3.5 en in tabel 3.5 is te zien dat voor elke conditie de testscore in de tijd afneemt. Tegelijkertijd wordt ook weergegeven dat op elk testmoment de conditie X2 (dissectietaak, benoemen en groepsbespreking) hoger scoort dan de conditie X1 (dissectietaak, tekenen en benoemen) met als laagste de controleconditie X0 (enkel de dissectietaak).

Resultaten over onderzoeksdesign 2

In tabel 3.6 is een overzicht gegeven van de gemiddelde scores op de natoets ($O_{2.1}$ in figuur 3.2) met de bijbehorende standaarddeviaties. De resultaten van de T-toets, in tabel 3.7, geven aan dat er geen significant verschil is tussen de experimentele condities, X2 ($M = 0,810$) en X1 ($M = 0,819$).

Tabel 3.6 *Gemiddelde testcores (M) en standaarddeviaties(SD) voor de experimentele condities op de natoets van alle studenten (N=140) die aan het practicum hebben deelgenomen*

Experimentele condities	M	SD
X1	0,819	0,095
X2	0,810	0,094

Noot: X1: experimentele conditie 1;: oorspronkelijke instructie (dissectietaak, tekenen en benoemen); X2: experimentele conditie 2;: aangepaste instructie (dissectietaak, benoemen en groepsbespreking).

Tabel 3.7 *T-toets met 2 onafhankelijke steekproeven voor de experimentele condities(X1 en X2)*

F	Significantieniveau ($p \leq 0,05$)	Std. Error difference	95% betrouwbaarheidsinterval van het verschil	
			beneden	Boven
0,009	0,926	0,0226	-0,0545	0,03564

De testen zijn ook hier na de natoets nog tweemaal aangeboden in twee retententietoetsen (O_{2.2} en O_{2.3} in figuur 3.4). De resultaten daarvan zijn te zien in tabel 3.8 waarin van de drie tests de gemiddelde testcores en de standaarddeviaties zijn weergegeven.

Tabel 3.8 *Gemiddelde testcores (M) en standaarddeviaties (SD) voor de experimentele condities op de natoets en retententietoetsen van alle studenten (N=74) die alle drie testen hebben gemaakt*

Experimentele condities		M	SD
X1	Natoets	0,837	0,072
	Retententietoets 1	0,806	0,116
	Retententietoets 2	0,775	0,111
X2	Natoets	0,838	0,097
	Retententietoets 1	0,813	0,105
	Retententietoets 2	0,739	0,092

Noot: X1: experimentele conditie 1;: oorspronkelijke instructie (dissectietaak, tekenen en benoemen); X2: experimentele conditie 2;: aangepaste instructie (dissectietaak, benoemen en groepsbespreking).

Uit de resultaten van de repeated measures analyse in tabel 3.9 blijkt dat er verschillen zijn in gemiddelde scores tussen de tijdstippen waarop de testen zijn afgenomen ($F = 27,851$, $p < 0,01$). Er geen significant verschil tussen de condities ($F = 3,089$, $p > 0,05$) en er is geen significant interactie-effect. Om een beeld te geven van de verschillen in tijdstippen voor de testafname zijn de gemiddelde scores per testtijdstip en conditie weergegeven in figuur 3.6.

Tabel 3.9 *Repeated measures analyse voor de experimentele condities op drie tijdstippen*

	Df	Variantie	F	Overschrijdingskans
Testtijdstippen	2	0,058	17,645	0,000
Interactie tijdstippen en condities	2	0,046	1,413	0,251
Conditie	1	0,002	0,097	0,758

Noot: Significantieniveau $p \leq 0,05$.

Noot: X1: experimentele conditie 1; oorspronkelijke instructie (dissectietaak, tekenen en benoemen); X2: experimentele conditie 2; aangepaste instructie (dissectietaak, benoemen en groepsbespreking).

Figuur 3.6 Gemiddelde testscores voor de experimentele condities op de natoets en retentientoetsen (N=74)

Betrouwbaarheid

Als maat voor de betrouwbaarheid van de natoets en retentientoetsen is gebruik gemaakt van Cronbach's alpha. De resultaten staan in tabel 3.10. Alpha geeft de ondergrens van de betrouwbaarheid weer in een schaal van 0 tot 1. Hoe hoger de alphawaarde, hoe betrouwbaarder de test. Als algemene vuistregel wordt een waarde van 0,7 toegepast als minimaal vereiste waarde. De toets zoals hier bestudeert functioneert als één van de toetsen in een module en is als zodanig geen 'high stakes' toets. Hierbij wordt de COTAN 2010 gevolgd waarbij een betrouwbaarheid van 0,7 als voldoende wordt beschouwd voor zo'n test.

Tabel 3.10 *Betrouwbaarheidsmetingen van het meetinstrument op drie meetmomenten*

Testen	Cronbach's alpha	Variantie	S.E. of Measurement
Natoets 1	0,6246	0,275	0,101
Retententietoets 1	0,7287	0,319	0,093
Retententietoets 2	0,6564	0,413	0,119

3.6 DISCUSSIE EN CONCLUSIE

Alle data en analyses zijn gebaseerd op scores op de meerkeuzetoets. Uit de betrouwbaarheidsanalyse blijkt dat de betrouwbaarheid in alle drie testen net onder of boven de ondergrens ligt van 0,7. De meest voor de hand liggende conclusie is dan ook dat de toets voor verbetering vatbaar is. Het is opvallend dat de alpha in de eerste retentietoets toeneemt om daarna weer af te nemen. Hier is geen toepasselijke verklaring voor. Behalve het te meten construct zijn er dus storende factoren die van invloed zijn en is een itemanalyse vereist waarbij items worden toegevoegd en 'slechte' items verwijderd. In het volgende hoofdstuk wordt dieper ingegaan op het te meten construct, de inrichting van de toets en de analyse van de papierentoets. Voor wat betreft de bruikbaarheid van de toets in het hier gepresenteerde experiment zijn de experimentele resultaten wel bruikbaar omdat het gaat om het vergelijken van instructies en niet om de constructie van de toets en diens effectiviteit. Aangenomen wordt dat fouten in de toets systematische meetfouten zijn die evenredig gelden voor beide instructies. De vergelijking tussen beide instructies is daarmee nog geldig.

De resultaten van de onderzoeksgroep die deelnam aan het experiment daalden fors vanaf de eerste afname tot en met de derde afname, evenals de groepsgrootte. Hiervoor zijn twee mogelijke verklaringen: (1) de studenten wisten niet van de testafnames en konden zich als zodanig niet hierop voorbereiden alsook zich niet opgeven voor deelname. (2) Doordat het onderwijs geen strikte aanwezigheidsplicht kent kon niet vooraf worden bepaald hoeveel studenten op de gekozen momenten aanwezig zouden zijn. Wel is getracht in te schatten op welk moment de meeste studenten aanwezig zouden zijn.

Uit de resultaten blijkt een klein maar significant verschil tussen de condities met taakgerichte opdrachten en de controle conditie. Dit betekent dat hypothese 1 mag worden geaccepteerd. Voor de experimentele conditie

dissectie+tekenen+benoemen was dit resultaat te verwachten daar uit het onderzoek van Lansing (1981) ook bleek dat tekenen positief bijdraagt aan het leerproces in vergelijking tot enkel observeren en uit het onderzoek van Dean en Kulhavy (1981) bleek dat benoemen ook positief bijdraagt aan het leerproces. Door de combinatie van deze twee taken lijkt het overbodig om een object meerdere malen te moeten tekenen voor een positief leereffect zoals in de literatuur werd gesuggereerd door Lansing (1984). Ook voor de conditie dissectie+ benoemen+feedback waren de onderzoeksresultaten te verwachten omdat benoemen een positief effect heeft (Dean & Kulhavy, 1981) evenals feedback (Wang, Haertel & Walberg, 1990, 1993; Gagné, 1985; Reigeluth, 1994; Merrill, 1983) in een groepsbespreking (Smit et al, 2009). Voor dit experiment kan geconcludeerd worden dat instructies met taakgerichte opdrachten positief bijdragen aan het visueel leerproces zoals dat in dit onderzoek is beschreven. Echter, uit de retentietoetsen blijkt dat dit verschil verdwenen is in een periode van drie weken. Hoe groot de invloed van de verschillende taakgerichte opdrachten en groepsfeedback afzonderlijk zijn is niet duidelijk omdat de opdrachten alleen in combinatie met elkaar zijn uitgevoerd.

Aangezien de experimentele condities X1 en X2 weinig verschillen tonen, kan geconcludeerd worden dat er in dit onderzoek geen verschil is gevonden tussen de twee instructievormen met opdrachten. Doordat dit verschil er niet is mag hypothese 2 worden geaccepteerd die stelt dat de aangepaste instructievorm minstens even effectief is als de bestaande instructievorm. De resultaten van de retentietoetsen geven aan dat de beklijving van kennis in een periode van drie weken voor beide condities even snel afneemt. In termen van Molenda (2009) zijn beide instructies even effectief.

Vervolgens bleek de aangepaste instructievorm in vergelijking tot de bestaande instructievorm op drie punten efficiënter, naar definitie van Molenda (2009). Allereerst is tekenen een arbeidsintensief proces dat de nodige tijd vereist waardoor de studenten continue onder tijdsdruk werkten om de tekenopdrachten te voldoen. De groep met de aangepaste instructie stond niet onder een vergelijkbare tijdsdruk en had voldoende tijd om met elkaar over de stof te discussiëren. Aan het eind van de dag waren studenten uit de aangepaste instructie gezamenlijk een uur eerder klaar dan was gepland terwijl een aantal studenten uit de bestaande instructie een vertraging hadden. Het resterende uur uit de aangepaste instructie zou in een later stadium kunnen worden benut voor

een daarop aansluitende toets. Doordat ze ook al feedback hadden gekregen over hun leerprestaties tijdens de discussie mag worden geconcludeerd dat de aangepaste instructie ook op dit aspect een efficiëntieslag heeft gemaakt. Vervolgens is de aangepaste instructie op een derde aspect efficiënter: te weten de reductie van de arbeidsintensieve evaluatie van de leerresultaten door het geautomatiseerd nakijken van een toets (140 stuks, een per student) in plaats van het nakijken van tekeningen (1400 stuks, 1 set van 10 per student). De correctieduur voor het geautomatiseerd nakijken van de toetsformulieren was een dagdeel. De ontwikkeling van de toets is een eenmalige investering. De correctieduur van een set tekeningen is ca 30 minuten wat voor 140 studenten resulteert in een totale correctieduur van 70 uren.

De onderzoeksresultaten geven weer dat het verantwoord wordt geacht om in het verdere verloop van het onderzoek de aangepaste instructievorm (dissectie, benoemen, groepsfeedback) toe te passen. Op grond van de resultaten in het onderzoek mag worden aangenomen dat de kwaliteit van de aangepaste instructievorm voldoende gelijkwaardig is aan de oorspronkelijke instructievorm om de oorspronkelijke instructievorm te vervangen. In combinatie met de tijdswinst, de al voorziene feedback aan de student en de aansluiting op de te ontwikkelen CBT toets is het gerechtvaardigd om de aangepaste instructie in het onderwijs toe te passen. Daarmee is een goede basis gecreëerd om het onderzoek te vervolgen en dieper in te gaan op het te leren construct en de inrichting van de toets als testbed voor verdere experimenten ter beantwoording van de in de hoofdstukken 1 en 2 beschreven meer specifieke onderzoeksvragen.

HOOFDSTUK 4

Constructie van de toets

In het hier gepresenteerde hoofdstuk wordt de papieren toets, ontworpen ten behoeve van het experiment beschreven in hoofdstuk 3, geëvalueerd en vervolgens gebruikt voor de doorontwikkeling naar een eerste versie van de voor de praktijk beoogde CBT toets. Het betreft een versie in vier varianten die met continue betrokkenheid van inhoudelijk experts wordt ontwikkeld. Het ontwikkelen van deze toets vormt de tweede ontwerpcyclus in de ontwerpgerichte aanpak van dit project. De vier varianten dienen voor experimenten die de in hoofdstuk 2 genoemde onderzoeksvragen moeten beantwoorden ten aanzien van de meest geschikte visualisatievormen voor de items van de uiteindelijke toets. De experimenten vormen een derde en vierde cyclus in de vormgeving van de toets en worden beschreven in de hoofdstukken 5 en 6.

4.1 ANALYSE VAN DE 1^{STE} TOETS: PAPIERENVERSIE

De in hoofdstuk 3 pragmatisch ontwikkelde eerste toetsversie komt deels tegemoet aan de in hoofdstuk 1 genoemde eisen voor een te construeren toets, (a) aansluiting op het leerproces, (b) objectieve meting en (c) snelle feedback. De toets is in samenspraak met de vakdocenten geconstrueerd na observatie van de oorspronkelijke instructie, bestudering van het instructiemateriaal en gesprekken met de vakdocenten en studentassistenten. De toets is getest in het vorige hoofdstuk en uit de resultaten blijkt dat onvoldoende voldaan is aan de minimaal vereiste Cronbach's waarde van 0,7 (COTAN, 2010). Voor de test en twee retentietoetsen zijn de waarden respectievelijk 0,62; 0,72 en 0,65. Geconcludeerd is in hoofdstuk 3 dat de alphawaarde een ondergrens aangeeft voor de betrouwbaarheid en met deze resultaten rondom de minimale grens is de conclusie dat de toets voor verbetering vatbaar is. Met dit resultaat is het ook aannemelijk dat er naast het te meten construct storende factoren aanwezig zijn die de kwaliteit van de toets beïnvloeden.

In deze paragraaf wordt, voor de verdere ontwikkeling van de toets naar een CBT-toets, specifiek ingegaan op de kwaliteit van de toets door eerst de toetsitems te analyseren en vervolgens de toets met vakdocenten te evalueren.

Voor de analyse van de toetsitems staan in tabel 4.1 voor elk item weergegeven het gemiddelde, de variantie en de punt-biseriële correlatie ofwel de item-testcorrelatie voor dichotome items. De test zelf staat weergegeven in de bijlagen die horen bij hoofdstuk 3.. Het gemiddelde is indicatief voor de moeilijkheid van het item en de variantie is een indicatie van de bijdrage van het item aan de betrouwbaarheid. Hiermee kunnen zwakke items geïdentificeerd worden. De punt-biseriële correlatie onderscheidt de laag- en hogescorende items en is daarmee indicatief voor de itemdiscriminatie, en is eveneens een indicatie van de bijdrage van het item aan de betrouwbaarheid van de toets.

Tabel 4.1 *Itemanalyse papierentoets*

Toets itemnr (N _{items} =41)	Gemiddelde (p-waarde)	Variantie	Punt-biseriële correlatie	Toets itemnr (N _{items} =41)	Gemiddelde (p-waarde)	Variantie	Punt-biseriële correlatie
1	0,74	0,20	0,350**	(21)	0,98	0,02	0,105
2	0,51	0,25	0,253**	22	0,73	0,20	0,429**
3	0,77	0,18	0,352**	23	0,81	0,16	0,203*
(4) ¹	0,91	0,08	0,255**	24	0,72	0,20	0,267**
(5)	0,93	0,07	0,122	(25)	0,82	0,15	0,093
(6)	0,80	0,16	0,198*	26	0,78	0,17	0,365**
(7)	0,89	0,10	0,190*	(27)	0,93	0,07	0,162
8	0,55	0,25	0,275**	(28)	0,29	0,21	0,052
9	0,85	0,13	0,347**	(29)	0,98	0,02	0,208*
(10)	0,72	0,20	0,191*	(30)	0,92	0,07	0,108
(11)	0,89	0,10	0,326**	31	0,46	0,25	0,233**
(12)	0,99	0,01	0,127	(32)	0,96	0,04	0,063
13	0,66	0,23	0,501**	33	0,56	0,25	0,325**
14	0,87	0,11	0,283**	34	0,82	0,15	0,291**
15	0,66	0,23	0,371**	(35)	0,98	0,02	0,193*
(16)	0,90	0,09	0,240**	(36)	0,97	0,03	0,188*
17	0,53	0,25	0,277**	(37)	0,99	0,01	-0,131
(18)	0,95	0,05	0,290**	(38)	0,75	0,19	0,168*
(19)	0,76	0,19	0,187*	(39)	0,89	0,10	0,212*
(20)	0,83	0,14	0,180*	(40)	0,97	0,03	0,168*
				41	0,71	0,21	0,300**
Test gemiddelde	0,7981			N _{items} =41			**= sign. 0,01
Cronbach alpha (α)	0,625			N _{stud} =140			*= sign. 0,05

Noot: Tussen haakjes de zwakke items die weinig toevoegen aan de toets, norm: var ≤ 0,1 of PB ≤ 0,2.

Uit bovenstaand overzicht blijkt dat

1. De test als geheel een lage betrouwbaarheid heeft. Een analyse waarbij items werden verwijderd die een negatieve invloed hadden op de betrouwbaarheid niet leidde tot geschikte resultaten met een alphawaarde groter dan 0,7. Ook een categorisatie naar type vraag (tekstueel, met diagram, met schematische weergave, zwart-wit foto & kleurenfoto) leverde per categorie geen alphawaarde groter dan 0,7 op.
2. Er items aanwezig zijn die weinig bijdragen aan het onderscheidend vermogen dat van een toets wordt verwacht. Op basis van afwegingen tussen de gemiddelde, itemvariantie en punt-biseriële correlatie zijn die items geselecteerd die weinig toevoegen aan de betrouwbaarheid van de toets. Het zijn zwakke items (in tabel 4.1 aangegeven tussen haakjes) die moeten worden gewijzigd of verwijderd op grond van een lage variantie (kleiner dan of gelijk aan 0.1) of een kleine punt-biseriële correlatie (kleiner dan of gelijk aan 0.2). De zwakke items ($N_{\text{zwakke items}} = 24$, ca 59% van het totaal $N=41$) zijn vooral items die heel hoog scoren. In bijlage 4.1 (bijlagen horende bij hoofdstuk 4), zijn de items weergegeven conform het itemnummer. Van de items met een hoge score bestaat het overgrote deel uit items met een visualisatie. Dit is te verklaren vanuit Goldstein (1985) en Levie (1987) die aangeven dat bij items met visualisaties rekening gehouden moet worden met het buitengewone menselijk vermogen om plaatjes te herkennen die eerder zijn gezien. Levie (1987) stelt dat het vermogen tot herkenning zo goed is dat onderzoekers zich moeten richten op het voorkomen van plafondeffecten bij de ontwikkeling van herkenningstesten.
3. Er in de toets twee itemgroepen zijn die weinig toevoegen aan de toets. Items 12 en 36 zijn van hetzelfde type itemformat en visualisatie, bevragen hetzelfde onderwerp (ventrale spieren) en kunnen worden beschouwd als een groep van hetzelfde thema. Items 21 en 35 zijn ook van hetzelfde type itemformat en visualisatie, bevragen hetzelfde onderwerp (mannelijke geslachtsorganen) en kunnen ook beschouwd worden als een groep van hetzelfde thema. Op al deze items is de score dusdanig hoog en wijzen de bijbehorende gegevens uit dat de items niet onderscheidend zijn. Daaruit kan niet geconcludeerd worden dat dit probleem uitsluitend wordt veroorzaakt door de visualisatievorm omdat andere itemtypen met vergelijkbare visualisatievorm wel onderscheidend zijn.

Over de zwakke items die moeten worden gewijzigd of verwijderd worden in de tweede ontwerpcyclus beslist in de doorontwikkeling naar een CBT-toets.

Na de psychometrische analyse is de toets met vakdocenten geëvalueerd. Deze evaluatie leverde de volgende resultaten op:

1. Er was tevredenheid over het feit dat de papierentoets, in vergelijking tot het beoordelen van tekeningen, een objectieve beoordelingsmethode bleek die in een veel korter tijdsbestek de toetskandidaten van feedback voorzag. Doorontwikkeling van de toets naar een CBT-toets biedt door automatische scoring de mogelijkheid in een nog korter tijdsbestek de toetskandidaten van feedback te voorzien.
2. Volgens de vakdocenten benaderden de toetsitems in de papierenversie de leersituatie nog onvoldoende. De toets zou nog te weinig items hebben waarin gebruik werd gemaakt van gedetailleerde representaties: meer specifiek representaties die in detail het te bestuderen object benaderen zoals foto's en gedetailleerde tekeningen.
3. De vakdocenten gaven ook aan dat de papierentoets beperkt bleek in de mogelijkheden om met visuele representaties meer open vraagtypen te ontwikkelen zoals CFR-items. In een papieren toets zijn de meeste opties die CFR-items bieden, zoals het puzzelen met gevisualiseerde objectonderdelen, praktisch gezien moeilijk uitvoerbaar.
4. De vakdocenten gaven als vierde punt aan te twijfelen aan de constructdefinitie die eerder met hen was overeengekomen. Deze zou wel dekkend zijn voor de tot nu toe opgestelde leerdoelen die jaren zijn toegepast maar bij nader inzien onvoldoende dekkend zijn voor wat studenten daadwerkelijk leren tijdens het practicum (het geleerde construct). Dit onderwerp kwam eerder ter sprake naar aanleiding van de analyse van het vak, weergegeven in tabel 3.1 (hoofdstuk 3). In combinatie met de eerste ontwerpcyclus en de vergelijking van de instructies, kwamen de vakdocenten tot de conclusie dat tijdens de instructie veel informatie wordt gegeven over de diverse anatomische structuren. Dat gaat gepaard met hun functionaliteit of functionele relatie tot andere structuren wat niet is weergegeven in de toets. De docenten gaven aan dat dit reden is om te overwegen de constructdefinitie op dit punt aan te passen.

Het resultaat van de toetsanalyse is verklaarbaar vanuit de begripsvaliditeit van de toets (Van den Brink & Mellenbergh, 2003). Een onderdeel van begripsvaliditeit is de dekking van het inhoudelijk domein. Tijdens en na samenstelling van de toets dient continue te worden gecontroleerd of het gehele construct inhoudelijk gedekt is in de toets. Daarvoor zijn er voornamelijk twee

factoren die van invloed zijn, te weten (a) construct-onderrepresentatie en (b) construct-irrelevantie. Als er sprake is van een construct waarin niet alle kenmerken gedekt zijn wordt er gesproken van construct-onderrepresentatie (Messick, 1993; Cambell & Fiske, 1959). Dat is het verschijnsel waarbij meetfouten worden veroorzaakt doordat niet alle kenmerken van het construct vertegenwoordigd zijn in de toets en de toets eigenlijk 'te weinig' meet. Het tegenovergestelde is construct-irrelevantie (Messick, 1993; Cambell & Fiske, 1959): de test 'meet teveel'. Hierbij wordt de kwaliteit van de toets bedreigt doordat er kenmerken worden gemeten die niet gerelateerd zijn aan het gedefinieerde construct. Er zijn twee vormen van construct-irrelevantie namelijk construct-irrelevante gemakkelijkerheid en construct-irrelevante moeilijkerheid. 'Construct-irrelevante gemakkelijkerheid' is het verschijnsel waarbij de test zo gemakkelijk is dat een hoge toetsscore behaald kan worden zonder dat men de leerstof beheerst. Kenmerkend is dat er een plafondeffect van de toetscores optreedt. Zo blijkt ook uit de analyse van de papierentoets waarin ca 30% items uit de test te gemakkelijk blijken te zijn en het toetsgemiddelde een waarde heeft van 0,798. De conclusie dat er sprake is van construct irrelevante gemakkelijkerheid is daarmee plausibel. De analyse biedt geen uitsluitsel of er sprake is van construct-onderrepresentatie. In de evaluatie met de docenten hebben de docenten echter aangegeven dat de toets op zichzelf dekkend is voor het gedefinieerde construct, maar bij nadere bestudering niet voor het daadwerkelijk geleerde construct waardoor dient te worden geëvalueerd of de gekozen constructdefinitie de meest wenselijke is.

Bovenstaande argumenten hebben ertoe geleid dat in samenspraak met de vakdocenten is besloten de doorontwikkeling van de papierentoets naar een CBT-toets in deze ontwerpcyclus op te pakken door het stappenplan van Oosterveld en Vorst (1998) nog eens te doorlopen met als start de evaluatie van de constructdefinitie en het kennisdomein in samenspraak met de stafleden van de faculteit. Dit stappenplan is daardoor ietwat uitgebreid en omvat de volgende onderdelen: (1) het evalueren en (her)definiëren van het construct, (2) het omschrijven van het te toetsen kennisdomein, (3) het specificeren van item- en testkenmerken, (4) het genereren van een verzameling items evenals het samenstellen van de toets, (5) het relateren van de toets aan de onderzoeksvragen en (6) het beschrijven van de psychometrische analyses voor de evaluatie van de toets alsook de toetsing van de onderzoeksvragen. In de volgende subparagrafen wordt de ontwikkeling van de toets volgens deze stappen uiteengezet.

4.2 (HER)DEFINIËRING VAN HET CONSTRUCT VOOR DIT ONDERZOEK

In deze paragraaf wordt het construct geëvalueerd en geherdefinieerd. In hoofdstuk 1 en 3 is het construct omschreven als visuele kennis over een concreet object, de rat, in het bijzonder (a) de visuele kenmerken van anatomische structuren zoals vorm, grootte, textuur en kleur, (b) de ruimtelijke relaties tussen de anatomische structuren met de relatieve grootte van elke structuur in relatie tot andere structuren, (c) de verbale labels en beschrijvingen van de structuren. Het omschreven construct is in overeenstemming met de kennisvorm die Hegarty (2005) bestempelt als de statische configuratie van een systeem. Nu de evaluatie met de docenten geleid heeft tot de vraag of deze constructdefinitie moet worden aangepast, is als volgt gewerkt.

Toegepaste werkwijze en methode

Met inhoudelijke experts, zijnde de docenten en de stafleden van de faculteit, is de definitie van het construct geëvalueerd om te beoordelen of de definitie van het construct in de huidige vorm volledig en dekkend is voor de kennis die wordt opgedaan over de anatomie van de rat. Een dergelijk evaluatie met inhoudelijke experts is verantwoord (Mc Kenney, Nieveen en van den Akker, 2006) waarbij gebruik wordt gemaakt van interviews met docenten en stafleden als zijnde domeinexperts die kunnen aangeven op welke wijze dit basisvak 'Anatomie van de rat' zich verhoudt tot de daaraan gerelateerde specifiekere vakken die de student later in de studie krijgt.

Toegepaste methode

De evaluatie heeft plaatsgevonden met semigestructureerde interviews waarin met een vragenlijst is nagegaan

- a. wat de doelstellingen zijn van de instructie,
- b. welke vaardigheden of kennis de student moet beheersen na het volgen van de instructie,
- c. waarop de student moet worden getoetst.

Ter voorbereiding kregen de respondenten twee weken vooraf de syllabus die hoort bij de instructie en was er een groepsbijeenkomst. Tijdens deze bijeenkomst werd door de onderzoeker verteld wat de doelstelling was van het interview en werd met elke respondent afzonderlijk een afspraak gemaakt. Er zijn in totaal 12

gestructureerde interviews afgenomen waarbij de vragen eerst in algemene zin werden gesteld over de gehele instructie en daarna meer toegespitst op de afzonderlijke instructiethema's. In tabel 4.2 staat een getalsmatig overzicht van de resultaten van de interviews over de instructie in zijn geheel.

Tabel 4.2 *Overzicht van de resultaten van de interviews (N=12)*

Vraagstelling	Antwoorden	Aantal personen (N=12)
Wat zouden de doelstellingen van het vak 'anatomie van de rat' moeten zijn?	Kennis en inzicht in de rat	4
	In groepsverband ontleden	1
	Inwendige habitus en bouw van een zoogdier	1
	Gevolgen van anatomische structuren/ waarom heeft iets een bepaalde structuur en welke functie is van belang?	1
	Kennis en inzicht in structuren door te weten hoe iets werkt en de functie daarbij kunnen bedenken	2
	Herkennen en benoemen van niet alleen organen maar ook systemen	2
Welke vaardigheden moet een student beheersen na het volgen van de instructie?	Het kunnen herkennen van organen door de vorm, textuur, kleur, grootte, verhoudingen	12
	Kunnen benoemen van structuren	12
	Benoemen in Latijn is niet belangrijk, dus:	• 3
	Benoemen in het Engels	▪ 2
	Benoemen met combinatie Engels en Latijn	▪ 1
	Benoemen in Latijn is belangrijk, dus:	• 9
	Benoemen in het Latijn	▪ 2
	Benoemen met combinatie Latijn en Engels	▪ 4
	Benoemen met combinatie Latijn en Nederlands	▪ 3
Waarop moet de student getoetst worden?	Het kennen van de posities van de organen ten opzichte van elkaar	8
	Het kennen van de functie van de verschillende structuren	8
	Herkenning van anatomische structuren	12
	Het leggen van relaties tussen de structuren, de naam en de functie van deze anatomische structuren	8
	Het benoemen van anatomische structuren	9

Verkrege resultaten

De resultaten van de interviews zijn door de onderzoeker geanalyseerd in samenwerking met de vakdocent als inhoudsdeskundige. Uit de analyses bleek dat de bestaande definitie van het construct inderdaad niet dekkend is voor het

geheel aan visuele kennis die wordt opgedaan uit de leerstof. Wat aan de constructdefinitie is toegevoegd is het leggen van relaties tussen anatomische structuren op basis van hun functie. Het enkel herkennen van anatomische structuren of kunnen benoemen van deze structuren is daarmee onvoldoende. Studenten moeten na de instructie in staat zijn om relaties te leggen tussen de structuren op basis van hun gemeenschappelijke of op elkaar afgestemde functies. Herkenbaar hierin zijn de kennisvormen causaal gedrag van het systeem en diens functies zoals eerder aangegeven door Hegarty (2005). Dit betekent kennis over de wijze waarop diverse structuren binnen het systeem elkaar beïnvloeden en welke functies daardoor worden vervuld.

De resultaten van de interviews hebben ertoe geleid dat de bestaande definitie van het construct is bijgesteld. Deze bijgestelde definitie van het construct omschrijft drie onderdelen die kunnen worden beschouwd als drie deelconstructen, zie figuur 4.1. Deze constructdefinitie zal in het verdere verloop van het onderzoek worden gehanteerd. Binnen de kaders van deze definitie kan invulling worden gegeven aan de inhoud van de toets.

Het te toetsen construct omvat kennis over de anatomische structuren van de rat.

Deze betreffen:

- de visuele kenmerken van deze structuren (vorm, grootte, kleur, textuur, spatiële verhoudingen)
- de verbale labels in het Latijn en beschrijvingen van deze structuren
- de relaties tussen deze structuren op basis van hun functionaliteit en causaal gedrag

Figuur 4.1 Definitie van het te toetsen construct

Bepalen van het te toetsen kennisdomein met experts

Om ervoor te zorgen dat de inhoud van de toets het kennisdomein van de anatomie van de rat dekt, is de test doorontwikkeld door de onderzoeker in samenwerking met de vakdocent. Daarbij is er gebruik gemaakt van (a) de definitie van het construct en (b) de vakinhoudelijke expertise van de andere docenten en stafleden over de leerstof in combinatie met (c) de syllabus waarin het kennisdomein is gespecificeerd.

Toegepaste methode

Ook hier is gebruik gemaakt van semigestructureerde interviews afgenomen bij de docenten en andere stafleden van de vakgroep waarbij is doorgevraagd over de leerstof.

Deze respondenten wordt per instructiethema gevraagd aan te geven

- a. wat een student moet leren uit de leerstof
- b. over welke onderdelen moet die student worden getoetst.

Er zijn in totaal 10 instructiethema's waarover bovenstaande vragen worden gesteld, te weten (a) habitus, (b) ventrale spieren, (c) situs viscerum buikholte, (d) uitgelegde darmtractus, (e) vrouwelijke geslachtsorganen, (f) mannelijke geslachtsorganen, (g) kop en hals ventraal, (h) situs viscerum borstholte, (i) grote bloedvaten en organen van de hals en (j) nieren, bijnieren en bloedvaten.

Verkregen resultaten

De interviews hebben de volgende resultaten opgeleverd:

1. Geen van de respondenten vindt het noodzakelijk het instructiethema habitus te verwerken in een toets omdat het voor de anatomie niet het meest relevante onderdeel is ter toetsing.
2. Onderwerpen die nauw samenhangen kunnen zowel in de instructie als in de toets worden samengevoegd.
  - instructiethema's (c) en (d) worden samengevoegd omdat beiden gaan over anatomische structuren betrokken bij de spijsvertering.
  - instructiethema's (e) en (f) gaan beiden over de geslachtsorganen en kunnen om deze reden worden samengevoegd.
  - instructiethema's (h) en (i) die de anatomische structuren in de borstholte en de hals behandelen met de verschillende bloedvaten worden samengevoegd.

Met de inhoudsdeskundigen is, met bovenstaande resultaten, het instructiemateriaal opnieuw bekeken. Besloten is ook het instructiethema (g), kop en hals ventraal, samen te voegen met instructiethema's (c) en (d) vanwege inhoudelijke overeenkomsten op het onderwerp spijsvertering. Het eindresultaat is dat de instructie 'Dissectie rat' bestaat uit de volgende instructiethema's die worden getoetst: (1) buikholte en spijsvertering, (2) spieren, (3) geslachtsorganen, (4) borstholte en (5) (bij)nieren en bloedvaten

De resultaten van de interviews worden gebruikt om nieuwe toetsitems te ontwikkelen en de items uit de eerste toetsversie (hoofdstuk 3), de papierenversie, te herzien en bij te stellen.

Specificatie van item- en testkenmerken

Voor het ontwikkelen of bijstellen van de itemvormen is gebruik gemaakt van de definitie uit figuur 4.1 in combinatie met de classificatie van representaties zoals weergegeven in tabel 2.1 (hoofdstuk 2). Deze definitie omschrijft drie deelconstructen die te vertalen zijn naar te toetsen taken.

Het eerste deelconstruct betreft het beheersen van kennis over de visuele kenmerken van anatomische structuren zoals vorm, grootte, kleur, textuur en verhoudingen. Uit de interviews blijkt dat domeinexperts belang hechten aan het toetsen van de mate van herkenning van anatomische structuren. In de literatuur over visuele toetsen (Dwyer, 1978) werd in het PSE-programma gebruik gemaakt van meerkeuzevragen voor herkenning. Ook in het toetsprogramma Ratview (De Graaff en Verhagen, 1995), dat gaat over het instructiethema “mannelijke geslachtsorganen van de rat”, wordt getoetst op herkenning waarbij de student uit een aantal anatomische structuren de gevraagde structuren kan kiezen. Gekozen is om op basis van deze informatie het construct over het herkennen van de visuele kenmerken van een structuur te operationaliseren door een meerkeuzevraag. Ook hier moet rekening worden gehouden met het buitengewone menselijk vermogen om plaatjes te herkennen die eerder zijn gezien (Levie, 1987; Goldstein, 1975). Zowel in het PSE-programma als in Ratview is ervoor gekozen om de vraag verbaal op te stellen en de antwoordalternatieven in de vorm van plaatjes over anatomische structuren weer te geven. Deze structuren bevatten alle typen visuele kenmerken die staan gespecificeerd in het construct. In samenspraak met de vakdocent is ervoor gekozen om het aantal antwoordalternatieven in elke meerkeuzevraag te beperken tot vier waarbij de kandidaat in de meeste meerkeuzevragen moet letten op twee visuele kenmerken. Dit betekent voor de antwoordalternatieven dat in het correcte antwoord beide kenmerken correct zijn, in twee afleiders is één kenmerk correct en één fout en in de derde afleider zijn beide kenmerken fout. Een voorbeeld van een meerkeuzevraag staat weergegeven in figuur 4.2. Dit item toetst de correcte oriëntatie en de correcte ingang van de rattenmaag. Dit type bevat naar aanleiding van tabel 2.1 alsook figuren 2.1 tot en met 2.3 (hoofdstuk 2) de categorie realistische representaties van (deel)objecten, al of niet getoond binnen het geheel van de context van het studieobject. Ten aanzien van het toetstype is gekozen voor toetstypen 2 en 4 in de categorisatie van Martinez (1998), tabel 2.2 (hoofdstuk 2) en tabel 2.3, waarin de taak behelst dat een beeld van een (deel)object wordt geëvalueerd op objectkenmerken.

Figuur 4.2 Voorbeeld van een itemformat waarin een anatomisch object moet worden herkend met als vraag: welk plaatje is de juiste weergave van een ventraal aanzicht van een rattenmaag?

Het tweede deelconstruct uit de definitie (figuur 4.1) is het kennen van de verbale labels en beschrijvingen van de anatomische structuren. Uit tabel 4.2 blijkt dat 9 van de 12 respondenten vindt dat er moet worden getoetst op het benoemen van anatomische structuren met de Latijnse naamgeving al dan niet in combinatie met de naamgeving in een andere taal zoals Engels of Nederlands. De Latijnse naamgeving is van belang omdat deze in de meeste gevallen weergeeft wat de functie van de betreffende structuur is. Om die reden is in samenspraak met de vakdocent gekozen voor het benoemen van de structuren met enkel de Latijnse terminologie. In de literatuur over visuele toetsen wordt ook aandacht besteed aan items waarin structuren worden benoemd. Het PSE-programma (Dwyer, 1978) bevat een toetsonderdeel dat de terminologietoets wordt genoemd en daarin krijgen de kandidaten een meerkeuzevraag over bepaalde feiten, benamingen of definities van structuren. Ook het programma Ratview (De Graaff en Verhagen, 1995) bevat een toetsonderdeel waarin studenten een anatomische structuur moeten benoemen. In dit onderdeel moeten studenten de correcte benaming van een structuur intypen en wordt ook gelet op eventuele schrijffouten.

Een voorbeeld van een vraag met een benoemingstaak staat weergegeven in figuur 4.3. Net als het vorige ontworpen itemformat bevat deze naar aanleiding van tabel 2.1 alsook figuren 2.1 tot en met 2.3 (hoofdstuk 2) de categorie

realistische representaties van (deel)objecten al of niet getoond binnen het geheel van de context van het studieobject. Ten aanzien van het toetstype is gekozen voor toetstype 2 in de categorisatie van Martinez (1998), tabel 2.2 (hoofdstuk 2) en tabel 2.3, waarin de taak behelst dat een beeld van een (deel)object wordt geëvalueerd en vervolgens benoemd.

Figuur 4.3 Voorbeeld van een itemformat over het benoemen van een anatomische structuur met als opdracht: benoem het aangegeven onderdeel van de rattenmaag.

In deze benoemingstaak krijgt de toetskandidaat een anatomische structuur te zien waarbij met een pijl is aangegeven welk onderdeel moet worden benoemd. Aan de pijl is een invulblok gekoppeld waarin de student de naam van de structuur kan invullen. Tegelijkertijd verschijnt er een lijst met de Latijnse namen van alle anatomische structuren die zijn behandeld tijdens de instructie. In samenspraak met de vakdocent staan in de lijst zowel correct als foutief geschreven Latijnse benamingen.

Het derde deelconstruct in figuur 4.1 is het leggen van relaties tussen anatomische structuren op basis van hun functionaliteit. Deze is geformuleerd naar aanleiding van de resultaten uit tabel 4.1 waaruit blijkt dat verschillende respondenten kennis over de functionaliteit van een anatomische structuur in relatie tot andere structuren belangrijk vinden. Acht van de twaalf respondenten geven aan dat er getoetst moet worden op relaties tussen structuren, de naam en de functie van deze anatomische structuren. Het PSE-programma (Dwyer, 1978) bevat geen

deeltoets die gericht is op het leggen van relaties tussen structuren op basis van hun functionaliteit. Wel bevat het programma een begripelijkheidstest waarin met behulp van verbale meerkeuzeitems naar bepaalde activiteiten van een anatomische structuur wordt gevraagd. In het programma Ratview wordt getoetst op de onderlinge relaties van de verschillende structuren en zo ook impliciet op hun functies door een puzzelelement in het item waarbij de diverse onderdelen met elkaar worden verbonden. De achterliggende redenering is dat als de toetskandidaat de correcte relaties legt tussen structuren, deze dat vooral doet vanuit de functionaliteit van het systeem. Deze systematiek is verwerkt in de operationalisatie van dit deelconstruct naar een toetstaak. Daarbij is gekozen om de toetskandidaat structuren met elkaar te laten verbinden als deze functioneel aan elkaar zijn gerelateerd. Voor deze toetstaak zijn CFR-items uitermate geschikt. Op basis van de in hoofdstuk 2 weergegeven taxonomie van CFR-items (Martinez, 1998) in relatie tot de leerdoelen, tabel 2.2 en tabel 2.3, is in samenspraak met de vakdocent gekozen voor de itemtypen 3, 5 en 6. Volgens deze itemformats krijgen de toetskandidaten de taken een relatie tussen structuren aan te geven op basis van categorische, ordinale of functionele relaties of een richting of pad tussen objecten aan te geven of objecten aan elkaar te koppelen in een bepaalde topografie (Martinez, 1998). Deze itemformats lijken het meest geschikt voor het operationaliseren van het derde deel van het construct. In figuur 4.4 is een voorbeeld van een item met een relatietaak weergegeven. In de stam van het item wordt aan de toetskandidaat gevraagd om een relatie tussen anatomische structuren weer te geven op basis van de erbij behorende functies. Op de structuren worden plekken gemarkeerd met groene punten. Met behulp van deze punten kunnen deze structuren met elkaar worden verbonden door objecten te verslepen naar de goede plekken of door het trekken van richtinggevendende pijlen ingeval naar de stroomrichting wordt gevraagd. Voor de representaties in deze toetstaak wordt gebruik gemaakt van de categorie diagrammen zoals weergegeven in tabel 2.1 (hoofdstuk 2).

Figuur 4.4 Voorbeeld van een itemformat over een functionele relatie

Het genereren van items en samenstelling van de toets

De volgende stap in het ontwikkelproces is het opstellen van toetsspecificaties en de daadwerkelijke constructie van de toets.

Voor dit onderdeel wordt gebruik gemaakt van (1) de aangepaste definitie van het construct dat dekkend is voor de te toetsen eigenschap en (2) de te toetsen leerstof.

De aangepaste definitie van het te toetsen construct omvat drie specifiekere deelconstructen die een bijdrage leveren. Deze dragen bij aan het opstellen van toetsspecificaties en worden geoperationaliseerd naar te toetsen taken met daaraan gekoppelde itemformats.

Opstellen van de vragen

De methode die wordt gehanteerd is hetzelfde als in de ontwikkeling van de eerste versie. In samenspraak met de vakdocenten worden de itemformats gevuld met de domeinspecificaties van het kennisdomein van de anatomie van de rat. Daarbij worden de items uit de papierenversie meegenomen in dit proces waarbij de geschikt bevonden items worden geselecteerd en aangepast. Om de toets inhoudelijk dekkend te laten zijn, zijn er eerst verbale vraagstellingen opgesteld die steeds de stam zijn van elke vraag. Het eindresultaat is een toets die bestaat

uit 55 items over drie deoltoetsen en inhoudelijk dekkend is voor het algehele construct en het kennisdomein. Tabel 4.3 geeft een overzicht van de aldus gegenereerde vraagstellingen.

Tabel 4.3 *Overzicht van de vragen in de toets, de instructiethema's, type itemformat en de vraagstellingen*

Vraag	Thema*	Type itemformat	Vraagstelling
1	1 (C/D/G)	Herkennen	Welk plaatje is de juiste weergave van het ventraal aanzicht van de rattemaag?
2		Benoemen	Benoem het aangegeven onderdeel van de rattemaag.
3		Herkennen	Welke klier is een weergave van de glandula submandibularis ?
4		Relatie	Geef door het verbinden van de verschillende organen de richting van het voedseltransport weer vanaf de maag, uitgaande van een duodenum van 12 cm .
5		Relatie	Geef door het verbinden van de verschillende organen de richting van het voedseltransport weer vanaf het coecum, uitgaande van een duodenum van 12 cm
6		benoemen	Benoem het aangegeven onderdeel van de rattemaag.
7		Herkennen	Welke klier is een weergave van de glandula lacrimalis ?
8		Herkennen	Welk plaatje is de juiste weergave van het duodenum?
9		Herkennen	Welk plaatje is een juiste weergave van het spijsverteringskanaal van maag tot en met rectum (het groen omkaderd deel is het duodenum met een lengte van 12 cm)?
10		Relatie	In het onderstaande schema zijn enkele anatomische structuren weergegeven of benoemd die bijdragen aan de afbraak van voedsel en aan de opname van stoffen in het bloed. Verbind deze structuren met elkaar zoals jij denkt dat ze in functioneel verband met elkaar staan.
11	2 (B)	benoemen	Benoem het aangegeven onderdeel van de rattemaag.
12		Herkennen	Welk deel van de kop is een weergave van de musculus masseter ?
13		Herkennen	Welk deel van de spieren is een weergave van de musculus deltoïdeus ?
14		Herkennen	Welk deel van de spieren is een weergave van de musculus rectus abdominus ?
15		Herkennen	Welk deel van de spieren is een weergave van de musculus rectus abdominus ?
16		Herkennen	Welk deel van de spieren is een weergave van de musculus obliquus abdominus externus ?
17		Herkennen	Welk deel van de spieren is een weergave van de musculi intercostales externi ?

Tabel 4.3 *Overzicht van de vragen in de toets, de instructiethema's, type itemformat en de vraagstellingen*

Vraag	Thema*	Type itemformat	Vraagstelling
18		Herkennen	Welk deel van de spieren is een weergave van de musculus obliquus abdominus internus ?
19	3 (E/F)	Benoemen	Benoem het aangegeven anatomisch onderdeel.
20		Relatie	Geef in de onderstaande afbeeldingen het transport van de onbevruchte eikel weer.
21		Benoemen	De rechtertestis is afgebeeld. Benoem het aangegeven anatomisch onderdeel.
22		Benoemen	De rechtertestis is afgebeeld. Benoem het aangegeven anatomisch onderdeel.
23		Herkennen	Welk plaatje is een juiste weergave van de rechtertestis en diens afvoergang?
24		Benoemen	Benoem het aangegeven anatomisch onderdeel.
25		Benoemen	Benoem het aangegeven anatomisch onderdeel.
26		Relatie	Kies uit de onderdelen van het mannelijk urogenitaalstelsel (rechterdeel) de correcte afbeeldingen en geef hun functionele relaties aan.
27		Benoemen	Benoem de aangegeven klier van de rechter mannelijke geslachtsorganen.
28		Benoemen	Benoem de aangegeven klier van de rechter mannelijke geslachtsorganen.
29		Herkennen	Welk plaatje is een juiste weergave van het mannelijk urogenitaalstelsel (rechterdeel)?
30		Benoemen	Benoem het aangegeven anatomisch onderdeel.
31		Benoemen	Benoem het aangegeven anatomisch onderdeel.
32		Relatie	Geef in de onderstaande afbeelding het transport van de spermatozoïden weer.
33		Relatie	Geef in de onderstaande afbeeldingen het transport van de spermatozoïden weer.
34	4 (H/I)	Herkennen	Welk plaatje is een correcte weergave van de vertakkingen van de venae pulmonales naar de longlobben vanuit ventraal aanzicht?
35		Herkennen	Welk plaatje is een weergave van de grote lichaamsvenen in relatie tot het hart? (ventraal aanzicht)
36		Benoemen	Benoem het aangegeven anatomisch onderdeel.
37		Benoemen	Benoem het aangegeven anatomisch onderdeel.
38		Benoemen	Benoem het aangegeven anatomisch onderdeel.
39		Herkennen	Welk plaatje is een juiste weergave van de vertakkingen van de aortaboog? (ventraal aanzicht)
40		Benoemen	Benoem het aangegeven anatomisch onderdeel.
41		Benoemen	Benoem het aangegeven anatomisch onderdeel.
42		Benoemen	Benoem het aangegeven anatomisch onderdeel.

Tabel 4.3 *Overzicht van de vragen in de toets, de instructiethema's, type itemformat en de vraagstellingen*

Vraag	Thema*	Type itemformat	Vraagstelling
43		Benoemen	Benoem het aangegeven anatomisch onderdeel.
44		Benoemen	Benoem het aangegeven anatomisch onderdeel.
45		Relatie	Geef aan op welke delen van het hart de verschillende onderdelen zijn aangesloten.
46		Herkennen	Welk plaatje is een correcte weergave van het hart met aan- en afvoerende bloedvaten? (ventraal aanzicht)
47		Relatie	Geef van de bloedvatuiteinden die rechtstreeks verbonden zijn met de rechterventrikel van het hart d.m.v. een pijl de bloedstroomrichting aan. (ventraal aanzicht)
48		Relatie	Geef van de bloedvatuiteinden die rechtstreeks verbonden zijn met de linkeratrium van het hart d.m.v. een pijl de bloedstroomrichting aan. (ventraal aanzicht)
49		Relatie	Geef van de bloedvatuiteinden die rechtstreeks verbonden zijn met de rechteratrium van het hart d.m.v. een pijl de bloedstroomrichting aan. (ventraal aanzicht)
50		Relatie	Geef van de bloedvatuiteinden die rechtstreeks verbonden zijn met de linkerventrikel van het hart d.m.v. een pijl de bloedstroomrichting aan. (ventraal aanzicht)
51	5 (J)	Relatie	Rostraal zijn de bloedstroomrichtingen in de lichaamsslagader en de onderste holle ader gegeven. (ventraal aanzicht) Verbindt op de juiste wijze – denk aan de bloedstroomrichting – deze bloedvaten met de rechter nier en met de rechter achterpoot.
52		Relatie	Rostraal zijn de bloedstroomrichtingen in de lichaamsslagader en de onderste holle ader gegeven. (ventraal aanzicht) Geef de bloedstroomrichting aan van de arteria iliaca en de vena iliaca.
53		Benoemen	In het onderstaande plaatje wordt de stroomrichting van het bloed in de arteriën en venen aangegeven (rode pijlen). Benoem de met een ster aangegeven arteriën en venen.
54		Herkennen	Welk onderdeel van de nier is een weergave van de medulla ?
55		Herkennen	Welk onderdeel van de nier is een weergave van de papillus ?

Noot: *) B = ventrale spieren (thema 2); C = situs viscerum buikholte (thema 1); D = uitgelegde darmtractus (thema 1); E = vrouwelijke geslachtsorganen (thema 3); F = mannelijke geslachtsorganen (thema 3); G = kop en hals ventraal (thema 1); H = situs viscerum borstholte (thema 4); I = grote bloedvaten en organen van de hals (thema 4); J = nieren, bijnieren en bloedvaten (thema 5).

Profiel van de toets

Door de drie (deel)constructen te vertalen naar drie toetstaken zijn er ook drie (deel)testen geconstrueerd, namelijk (1) een toets met herkenningstaken (20 items), (2) een toets met benoemingstaken (21 items) en (3) een toets met functionele relatietaken (14 items). Deze (deel)testen samen resulteren in een totale toetsscore. Ook in het PSE-programma en in het Ratview-programma moest een totaalscore worden berekend doordat de toetsen uit verschillende onderdelen bestonden. In dit onderzoek wordt de totaalscore voor de toetskandidaat berekend door per homogene itemgroep van een toetstaak de gemiddelde score te bepalen en daarna deze scores samen te middelen. Dit betekent dat de toetstaken even zwaar wegen. Tijdens de interviews gaven de vakdocenten aan de drie constructen even belangrijk te vinden en even zwaar te willen laten meetellen.

Constructie van antwoordmogelijkheden in relatie tot de specifieke onderzoeksvragen

Het profiel van de toets zoals die nu ontwikkeld is, heeft tot doel te toetsen in hoeverre de lerende zich het geleerde construct heeft eigen gemaakt. De volgende stap is het construeren van de antwoordmogelijkheden van de testitems met geschikte visuele representaties voor het toetsen van de voor het hier gepresenteerde onderzoek geformuleerde specifieke onderzoeksvragen, die in hoofdstuk 2 beargumenteerd zijn:

- *Meet een toets met schema's hetzelfde visuele construct als een toets met foto's?*
- *Meet een toets met context hetzelfde visuele construct als een toets zonder context?*

De eerste specifieke onderzoeksvraag is geoperationaliseerd door visualisaties in fotovorm en in schematische vorm. De foto's visualiseren de dissectiesituatie het meest vanwege de kleuren en de vele details in vormen en texturen. De schematische tekeningen bevatten de contouren van objecten wat betekent dat alleen de kenmerkende vormen van de objecten zijn gevisualiseerd in zwarte pentekeningen. De tweede specifieke onderzoeksvraag is geoperationaliseerd door visualisaties waarin contextuele informatie is toegevoegd en visualisaties waarin de contextuele informatie afwezig is. In visualisaties zonder contextuele informatie zijn de objecten afzonderlijk vertoond. In visualisaties met contextuele informatie is naast de afzonderlijke objecten een overzicht van de buitenste contouren van de rat weergegeven met daarin een deel van de anatomische structuren die behoren tot de directe omgeving van het object dat in het betreffende item wordt bevraagd.

Deze operationalisaties van de onderzoeksvragen hebben geresulteerd in vier verschillende toetsen met de condities (a) kleurenfoto's met contextuele informatie (foto met context = FMC), (b) kleurenfoto's zonder contextuele informatie (foto zonder context = FZC), (c) schematische tekeningen met contextuele informatie (schema met context = SMC) en (d) schematische tekeningen zonder contextuele informatie (schema zonder context = SZC). Doordat deze vier varianten elk drie typen itemformats (operationalisaties van de constructen) bevatten zijn er in totaal twaalf visueel verschillende itemsformats geconstrueerd die staan weergegeven in figuur 4.5 en tabel 4.4. In tabel 4.4 staat verder het aantal items per itemformat per toetsvorm alsook het totaal aantal items per toetsvorm. Deze gegevens corresponderen met de gegevens in tabel 4.3 waar in een overzicht de vragen in de toets, de instructiethema's, type itemformat en de vraagstellingen staan weergegeven.

Tabel 4.4 *Overzicht van de verschillende itemformats en toetsvormen met het aantal items*

Toetsvormen (4)	→	foto met foto zonder schema schema			
		context (FMC)	context (FZC)	met context (SMC)	zonder context (SZC)
Itemformats (3) ↓	Herkenningstaak	20	20	20	20
	Benoemingstaak	21	21	21	21
	Functietaak	14	14	14	14
Totaal aantal items per toetsvorm		55	55	55	55

	Herkennen	Benoemen	Relaties
Schema zonder context (SZC)			
Schema met context (SMC)			
Foto zonder context (FZC)			
Foto met context (FMC)			

Figuur 4.5 De visualisatievormen van de 12 verschillende itemformats

In de bijlage 4.1, horende bij hoofdstuk 4, wordt een totaaloverzicht van de toets gepresenteerd waarbij de vier verschillende representatievormen worden weergegeven. Elke vraag geeft een van de visualisatievormen weer om een beeld te geven hoe de vier representatievormen er in de toets uitzien. De vragen 1 tot en met 14 geven de toetsvorm 'foto zonder context' (FZC) weer. De vragen 15 tot en met 28 zijn een weergave van de toetsvorm 'schema zonder context' (SZC). De vragen 24 tot en met 42 zijn een weergave van de toetsvorm 'foto met context' (FMC). Tot slot zijn de vragen 43 tot en met 55 een weergave van de toetsvorm 'schema met context' (SMC).

4.3 VERVOLG ONDERZOEK

Samenvattend is de papier-en-pen toets die de basis vormde voor de CBT-toets op verschillende aspecten gewijzigd tijdens de vertaalslag waardoor een compleet nieuwe toets is geconstrueerd. Alle wijzigingen zijn onder te brengen in drie categorieën:

Inhoudelijk zijn er 12 van de 41 items (ca. 30%) verwijderd waarvan er 7 definitief zijn verwijderd en de overige 5 zijn omgebouwd naar nieuwe vragen. Daarnaast zijn nog 21 vragen toegevoegd om het construct te dekken.

Visueel zijn alle visuele vormen opnieuw vormgegeven en beperkt tot het gebruik van schema's en foto's. De schema's en foto's zijn verder uitgewerkt naar weergaven met en zonder contextuele informatie. Dit resulteerde in nieuwe visualisaties in vergelijking tot de papier-en-pen toets die worden toegepast in vier parallelle varianten van dezelfde toets.

De *toetstaken* zijn uitgebreid met niet alleen herkenningstaken maar ook benoemingstaken waarin de anatomische onderdelen worden benoemd en functietaken waarin relaties worden gelegd tussen anatomische onderdelen.

Hoofdstuk 5 beschrijft het eerste experiment met de CBT-toets als volgende stap in de ontwikkeling van de instructie-toets combinatie waarmee de onderzoeksvraag kan worden beantwoord.

HOOFDSTUK 5

Visuele CBT-toets – 1e experiment

In dit hoofdstuk wordt een eerste experiment gepresenteerd met de ontwikkelde CBT toets waarin de ontwikkelde instructie-toetscombinatie centraal staat evenals de toegepaste representatievormen. Uit het experiment zal blijken dat de toets moet worden bijgesteld door de veelheid aan zwakke items. Het hoofdstuk beschrijft de analyse van de toets ter verbetering ervan en eindigt met een bijgestelde toets die geschikt is voor een volgend experiment.

5.1 INTRODUCTIE

Het in dit hoofdstuk gepresenteerde experiment is exploratief en wordt uitgevoerd met de aangepaste instructie uit hoofdstuk 3 en de in hoofdstuk 4 ontwikkelde CBT-toets. Het betreft de voortzetting van de prototypefase waarin langs twee onderzoekslijnen wordt gewerkt met als hoofddoel het evalueren van de ontwikkelde instructie-toetscombinatie. Tevens wordt met het experiment beoogd te onderzoeken wat de beste keuze is voor de vormgeving van de visuele representaties in de toets met betrekking tot de factoren realisme (schema's of foto's) en context (met of zonder).

De eerste onderzoekslijn is de beoordeling of de ontwikkelde instructie-toetscombinatie dekkend is voor het construct en of het construct in de verschillende toetsversies vergelijkbaar gerepresenteerd is. De toets wordt geconstrueerd vanuit toetsspecificaties met de veronderstelling dat op deze wijze het construct gedekt is. De vraag of de toets inderdaad dekkend is, is een inhoudelijke vraag. Als er echter veel items uitvallen omdat ze niet in de interne structuur passen of onbetrouwbaar zijn, wordt de dekking ondergraven.

De tweede onderzoekslijn is onderzoek naar de meest geschikte representatievorm(en) voor het toetsen van het construct zoals gespecificeerd in

de onderzoeksvragen (hoofdstuk 2) gelet op mate van realisme van representatievormen en de aan- of afwezigheid van contextuele informatie. Daartoe is de toets vertaald in vier representatievormen (hoofdstuk 4) die elk een experimentele conditie vormen: foto met context (FMC), foto zonder context (FZC), schema met context (SMC), schema zonder context (SZC). Deze worden gebruikt voor het beantwoorden van de volgende specifieke onderzoeksvragen:

- *Meet een toets met schema's hetzelfde visuele construct als een toets met foto's?*
- *Meet een toets met context hetzelfde visuele construct als een toets zonder context?*

Alvorens in te gaan op de uitvoering van het experiment, de verkregen resultaten en de daaruit voortvloeiende conclusies, wordt in de volgende paragraaf eerst ingegaan op de psychometrische analyses die worden toegepast om de onderzoeksvragen te beantwoorden.

5.2 PSYCHOMETRISCHE WERKWIJZE

Werkwijze ter toetsing van de kwaliteit van de toets

Er zijn vier scenario's als mogelijke resultaten van het experiment (a) het construct is goed gedekt en de vier visueel verschillende toetsversies meten hetzelfde, (b) het construct is goed gedekt en de verschillende toetsversies meten niet hetzelfde, (c) het construct is niet goed gedekt maar de verschillende toetsversies meten hetzelfde en (d) het construct is niet goed gedekt en de verschillende toetsversies meten niet hetzelfde.

Van belang in de bewijsvoering is de eerder al in hoofdstuk 4 omschreven begripsvaliditeit cq constructvaliditeit (Van den Brink & Mellenbergh, 2003). Daarbij is aan de orde in hoeverre het toetsresultaat geïnterpreteerd kan worden als indicatief voor het construct dat die toets beoogt te meten, en niet voor een ander construct, een deel ervan of juist een veel breder construct. Van den Brink en Mellenbergh (2003) geven aan dat steun voor constructvaliditeit moet worden afgeleid uit een combinatie van theoretische noties over een empirische toets en de daartoe vergaarde empirische gegevens. Dat betekent dat (1) het nomologisch netwerk van het construct moet worden gespecificeerd (2) de interne structuur dient te worden onderzocht, (3) de betrouwbaarheid van de toetsen moet worden onderzocht en (4) het inhoudsdomein voldoende dient te zijn gedekt.

Nomologisch netwerk

Een nomologisch netwerk is een representatie van een construct (concept) met andere verwante constructen, relaties met operationalisaties van die constructen en relaties tussen operationalisaties onderling (Cronbach & Meehl, 1955; De Groot, 1961). Dit netwerk kan bestaan uit hypothesen over empirische relaties. De toets moet daarvoor samenhangen met andere toetsen die hetzelfde construct meten. Ter indicatie hebben Campbell en Fiske (1959) een analysemodel ontwikkeld, de multitrek-multimethode-matrix, die een beeld biedt over de relatie tussen te toetsen deelconstructen (trekken) en de representatievormen/toetsversies (methoden). Toepassing van dit analysemodel vereist het correleren van de scores van de deelnemers tussen de methoden hetgeen gegeven de proefopzet in het hier gepresenteerde onderzoek niet mogelijk is daar deelnemers maar één toetsversie gepresenteerd kregen. Wel zijn de deelnemers in elke toetsversie getoetst op de drie deelconstructen waardoor correlaties tussen de toetstaken binnen elke toetsversie te schatten zijn. Dit levert vier correlatiematrices van correlatiestructuren binnen een toetsversie op waarna wordt nagegaan in hoeverre de correlatiestructuren binnen de toetsen met elkaar overeenkomen.

Interne structuur

De interne structuur van een toets betreft het aantal constructen dat een test meet. Een effectieve methode om dit in kaart te brengen is afkomstig uit de item-reponsetheorie (IRT) waarin een model wordt gespecificeerd naar aanleiding van het antwoord van de persoon op elk van de items. Uitgangspunt is de veronderstelling van de aanwezigheid van een niet direct observeerbare variabele, een latente variabele, die de respons op elk item afzonderlijk beïnvloedt. De antwoorden op de items worden gebruikt om te schatten welke positie de persoon inneemt op deze latente variabele. In de ideale situatie is deze variabele het te meten construct. Het doel van itemanalyse met een IRT-model is het aantonen dat de items van een bepaald construct inderdaad dat construct meten en geen andere constructen (Drenth & Sijtsma, 1990; Sijtsma & Molenaar, 2002). Er wordt gekeken naar de sterkte in samenhang tussen items door deze te schalen, op basis van de scores op een set items van elke persoon. Als de items die een construct zouden meten schaalbaar zijn op dezelfde schaal dan zijn ze unidimensionaal en mag worden geconcludeerd dat inderdaad hetzelfde construct gemeten wordt. Zijn er meerdere schalen dan betekent dat de aanwezigheid van meerdere factoren en worden er dus meer constructen gemeten dan één. In geval dit resultaat zich voordoet zal, afhankelijk van de verdeling in de verschillende schalen, bepaald moeten worden of hieruit aanwijzingen volgen

voor het bestaan van deelconstructen die overeenkomen met samenhangende onderdelen van de totale leerstof zoals die bijvoorbeeld in de omschrijving van de leerstof zijn genoemd in de vorm van vijf instructiethema's of dat er sprake is van ongewenste factoren die aanleiding zijn tot bijstelling van de toets.

Bij het schalen van items in het hier beschreven experiment, wordt gebruik gemaakt van een specifiek IRT-model, het Mokkenmodel, voor dichotome of gedichotomiseerde toetsitems. Dit model maakt het mogelijk om testgedrag te meten op een ordinale schaal. Er is gebruik gemaakt van een geautomatiseerd Mokkenanalyseprogramma, MSP, dat items schaalt waarbij met de schaalcoëfficiënt H (homogeniteit: zelfde aard of samenstelling) wordt bepaald of een item bij een Mokkenschaal hoort. Voor de Mokkenschaal geldt als norm een minimale waarde van 0,3. Een set van items met $H \leq 0,3$ vormt geen homogene Mokkenschaal. De normering voor de Mokkenschaal is: zwak bij $0,3 < H \leq 0,4$; gemiddeld bij $0,4 < H \leq 0,5$ en sterk bij $H > 0,5$.

In de praktijk is het gebruik van IRT-modellen vooral doeltreffend als het gaat om grote aantallen deelnemers. In het hier gepresenteerde onderzoek zijn de aantallen deelnemers niet substantieel groot. Om die reden kunnen de verwachtingen van het toepassen van de Mokkenanalyse niet te hoog gespannen zijn. De waarde van de Mokkenanalyse is hier gelegen in de verwachte geschiktheid om een trend zichtbaar te maken.

Betrouwbaarheid

Een indicatie voor de kwaliteit van de toets is het bepalen van de betrouwbaarheid. Daarvoor wordt een betrouwbaarheidsmeting uitgevoerd waarbij gebruik wordt gemaakt van Cronbachs alpha met als norm 0,7 of hoger (COTAN, 2010). Hoe meer items voldoende bijdragen aan een test, hoe betrouwbaarder de test. Met de gestelde norm is het een vereiste dat minimaal 70% van de variantie van de test is toe te wijzen aan de variantie van de ware score.

Inhoudelijk domein

De dekking van het inhoudelijk domein is in hoofdstuk 4 uitgebreid aan de orde geweest waarbij is ingegaan op de inhoudelijke validiteit van de toets. Deze is vooral geborgd door het intensief betrekken van domeinexperts bij de ontwikkeling van de toets. Psychometrische controle op de kwaliteit van de toets vindt achteraf plaats door, naast de betrouwbaarheid, een itemanalyse uit te voeren op de toetsresultaten.

Een itemanalyse biedt inzage in het functioneren van elk item afzonderlijk (itemgemiddelde en -variantie) en in de bijdrage van een item aan de betrouwbaarheid van de test (punt-biseriële itemrestcorrelatie). Met het itemgemiddelde wordt de itemmoeilijkheid bepaald die met een waarde tussen de 0 en 1 varieert van heel moeilijk tot heel gemakkelijk. Relevant daarbij is de itemvariantie want erg gemakkelijke en erg moeilijke items hebben een geringe variantie en zijn niet effectief in hun bijdrage aan de test. De punt-biseriële itemrestcorrelatie is indicatief voor de itemdiscriminatie die aangeeft in hoeverre een item onderscheid kan maken tussen personen die hoog scoren en personen die laag scoren. Daarbij wordt voor de punt-biseriële itemrestcorrelatie een ondergrens van 0,2 aangehouden. De gehanteerde norm voor zwakke items is een lage variantie ($\leq 0,1$) of een lage puntbiseriële correlatie ($PB \leq 0,2$).

Samengevat kunnen de resultaten, op basis van de hier beschreven analyses, iets zeggen over de mate waarin het construct al of niet goed is gedekt en in hoeverre de vier representatievormen van de toets al of niet hetzelfde construct meten. Ideaal is de situatie dat er geen items uitvallen en dat dus de items in de toets het te leren construct voldoende hebben gedekt. Met dat resultaat kan gekeken worden naar de geschikte representatievorm voor het toetsen van dat construct waarbij de deelconstructen tegen elkaar worden afgezet in elke toets voor de verschillende representatievormen. Er bestaat echter een reële kans dat de toets, in deze eerste versie, niet voldoende dekkend blijkt voor het te leren construct met als gevolg dat deze vervolgens moet worden aangepast ter verbetering alvorens de onderzoeksvragen in de tweede onderzoekslijn verder te beantwoorden. Dit resulteert dan in een bijstelling van de toets in het hier gepresenteerde onderzoek. De psychometrische analyses dienen hierover eerst uitsluitsel te bieden. Om die reden worden eerst het inhoudelijk domein bestudeerd, vervolgens de interne structuur en daarna het nomologisch netwerk.

Werkwijze ter toetsing van de verschillen tussen de verschillende toetsvarianten

In de tweede onderzoekslijn wordt onderzocht of er verschillen zijn tussen de toetsresultaten van de toetsvarianten foto met context (FMC), foto zonder context (FZC), schema met context (SMC), en schema zonder context (SZC). Mogelijke (hoofd-)effecten worden opgespoord met variantieanalyse waarbij gebruik wordt gemaakt van de gemiddelde scores op de vier toetsversies en -taken.

5.3 METHODEN

Onderzoeksdesign

In het experiment wordt de in hoofdstuk 4 geconstrueerde toets gebruikt in de vier combinaties van de variabelen realisme en context die de genoemde vier experimentele condities vormen. Het experiment vindt plaats volgens het 2*2 factoriële ontwerp dat is weergegeven in figuur 5.1.

		Context	
		Wel	Geen
Realisme	Kleurenfoto	FMC	FZC
	Schema	SMC	SZC

Figuur 5.1 2*2 factorieel ontwerp van de variabelen realisme en context

Het bijbehorende onderzoeksdesign is weergegeven in figuur 5.2.

R	X	FMC	<i>Noot:</i> R: at random toewijzing van studenten in de groepen; X: Instructie "Anatomie van de Rat"; FMC: Afname toetsversie kleurenfoto met context; FZC: Afname toetsversie kleurenfoto zonder context; SMC: Afname toetsversie schema met context; SZC: Afname toetsversie schema zonder context
	X	FZC	
	X	SMC	
	X	SZC	

Figuur 5.2 Onderzoeksdesign

Deelnemers

Het experiment is uitgevoerd in het tweede curriculumjaar van de opleiding farmacie bij de afdeling dierfysiologie. De onderzoeksgroep bestond uit 61 tweedejaars farmacistudenten waarvan $N_{FMC} = 16$, $N_{FZC} = 15$, $N_{SMC} = 15$ en $N_{SZC} = 15$. Deze studenten hadden eerder geen dissectie-instructie gevolgd en hadden geen dissectievaardigheden.

Procedure

De studenten werden eerst at random toegewezen aan één van de vier groepen en volgden daarna een instructie (X) om vervolgens aan één van de toetsen (FMC, FZC, SMC of SZC) deel te nemen. Studenten kregen voor de instructie een

introductie met informatie over de dissectie-instrumenten, de practicumprocedure en het veiligheidsreglement.

Voor alle groepen duurde de instructie twee dagen van elk acht uur. Per dag werd één instructiehelft doorlopen en elke dag werd beëindigd met een toetsafname over de instructieonderwerpen die dezelfde dag waren behandeld. De studenten werden voor het volgen van de instructie ingedeeld in instructiegroepjes van 10 à 11 personen. Elke groep kreeg een student-assistent toegewezen die als docent de instructie gaf. In totaal waren er zes student-assistenten die door de onderzoeker werden voorbereid op het experiment. Zij hadden geen voorkennis over de inhoud van de toetsen.

Voor de toets moesten de studenten zich verplaatsen naar een computerlokaal dat dienst deed als toetsruimte. Vooraf aan de toets werd een protocol doorlopen waarin de student mondeling werd ingeleid in de toets. Dit protocol omvatte het volgende:

- De studenten arriveren in het als toetsruimte gebruikte computerlokaal en moeten plaats nemen achter een computer,
- Klassikaal worden ze mondeling ingeleid in de toets met als onderdelen:
  - Het vermelden hoeveel items er in de toets zijn en welke type vragen men kan verwachten,
  - Het tonen van voorbeeldvragen. Per type vraag wordt een voorbeelditem getoond met gebruikmaking van een beamer. De drie typen items zijn herkennen, benoemen en het leggen van relaties,
- De studenten wordt verteld dat een ieder een eigen toets krijgt die enkel toegankelijk is met het eigen studentnummer,
- De studenten wordt verteld op welke wijze de beoordeling plaats vindt,
- De studenten krijgen kort de kans nog enkele vragen te stellen.
- Hierna loggen de studenten in met hun studentnummer om de toets te maken.

Dataverzameling

De toetsversies bestonden uit dezelfde items in dezelfde volgorde en waren identiek aan elkaar op de representatievormen na. Elke toets bestond uit 55 CBT-items waarvan 20 items met herkenningstaken, 21 items met benoemingstaken en 14 items met relatietaken (zie: bijlage hoofdstuk 4). In overeenstemming met de procedure van het experiment vond de dataverzameling plaats met twee deelttoetsen: één op elke dag over de op die dag behandelde thema's. Het hierboven beschreven toetsprotocol werd voor beide deelttoetsen op dezelfde wijze afgewikkeld.

In de data-analyse worden de data van de twee dagen geanalyseerd en met elkaar vergeleken. Als blijkt dat er geen significante verschillen zijn, zullen de data van elke persoon over die twee dagen worden samengevoegd. In hoofdstuk 3 is deze analyse ook uitgevoerd en daar bleken er geen significante verschillen te zijn in de prestatie die iemand leverde op de eerste en op de tweede dag. Het was toen verantwoord de data samen te voegen. Na de analyse zal blijken in hoeverre dat ook hier van toepassing is.

Data-analyse

Het significantieniveau van alle statistische toetsen is $p = 0,05$. De afhankelijke variabelen zijn (a) de scores op de gehele toets, (b) de scores op de herkenningstaak, (c) de scores op de benoemingstaak, en (d) de scores op de relatietaat. De onderzoeksresultaten zijn berekend voor elke toetsversie en toetstaak. Voor de bewijsvoering zijn Cronbach's alpha's, testgemiddelde, -variantie, itemgemiddelde, -variantie en punt-biseriële correlaties berekend. Bij scoring van de items met een relatietaat is rekening gehouden met het feit dat er meer verbindingen worden gemaakt. Daar is gekozen voor de proportie verbindingen die juist zijn geïdentificeerd zodat er opnieuw een getal ontstond tussen 0 en 1 als itemscore. Voorts zijn er Mokkenanalyses uitgevoerd, zijn de correlaties tussen de toetstaken voor elke toetsversie berekend en is nagegaan of de correlatiestructuur binnen de toetsen met elkaar overeenkomen. De methodiek rond de bewijsvoering is eerder in dit hoofdstuk beschreven.

Voor de analyse ten behoeve van de tweede onderzoekslijn naar de (hoofd)effecten van realisme en context is gebruik gemaakt van ANOVA.

5.4 RESULTATEN

De verschillen tussen de data vergaard tijdens de twee instructiedagen zijn bestudeerd door vergelijking van de testgemiddelden en varianties van alle testen gezamenlijk per dag en voor elke toetsversie afzonderlijk per dag. Middels T-toetsen zijn er geen significante verschillen gevonden tussen de twee instructiedagen. Om die reden kon zoals werd verwacht, worden besloten om voor de verdere verwerking van de resultaten de data van de deelttoetsen samen te voegen. Tabel 5.1a biedt een totaaloverzicht van de toetskwaliteit van de test: eerst zonder uitsplitsing en vervolgens uitgesplitst naar de toetsversies. Tabellen 5.1 b tot en met d zijn gedetailleerder voor elke toetstaak en voor elk item.

Tabel 5.1a *Overzicht analyse gehele test*

	Test algemeen (N= 61)			FMC (N= 16)			FZC (N= 15)			SMC (N= 15)			SZC (N= 15)		
	<i>Gem</i>	<i>Var</i>	<i>Cronbachs alpha</i>	<i>Gem</i>	<i>Var</i>	<i>Cronbachs alpha</i>	<i>Gem</i>	<i>Var</i>	<i>Cronbachs alpha</i>	<i>Gem</i>	<i>Var</i>	<i>Cronbachs alpha</i>	<i>Gem</i>	<i>Var</i>	<i>Cronbachs alpha</i>
T	0,64	0,019	0,85	0,60	0,015	0,82	0,63	0,023	0,88	0,66	0,025	0,89	0,66	0,015	0,82
H	0,80	0,013	0,51	0,79	0,008	0,14	0,79	0,017	0,63	0,77	0,019	0,64	0,84	0,008	0,44
B	0,63	0,040	0,78	0,58	0,046	0,81	0,63	0,040	0,80	0,67	0,047	0,82	0,63	0,031	0,72
R	0,49	0,031	0,72	0,43	0,017	0,49	0,48	0,039	0,79	0,53	0,036	0,77	0,52	0,031	0,73

Noot: T= Test (55 items); B= Benoemen (21 items); H= Herkennen (20 items); R= Relaties (14 items).

Tabel 5.1b Overzicht analyse toetstaak herkennen

HERKENNEN		Algemene Test (N= 61)			FMC (N= 16)			FZC (N=15)			SMC (N= 15)			SZC (N= 15)		
Itemnr	Type	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr
1	H	0,84	0,14	0,493**	0,81	0,16	0,645**	0,93	0,07	0,378	0,87	0,12	0,454	0,73	0,21	0,640**
3	H	0,92	0,07	0,254*	0,88	0,12	0,284	0,87	0,12	0,252	0,93	0,07	0,266	1	0	
7	H	0,9	0,09	0,316*	0,88	0,12	0,184	0,93	0,07	0,378	0,8	0,17	0,483	1	0	
8	H	0,44	0,25	0,183	0,5	0,27	0,205	0,6	0,26	0,032	0,2	0,17	0,383	0,47	0,27	0,319
9	H	0,72	0,2	0,162	0,63	0,25	0,23	0,4	0,26	0,041	0,93	0,07	0,266	0,93	0,07	-0,032
12	H	0,88	0,1	0,434**	0,75	0,2	0,560*	0,93	0,07	0,511	0,87	0,12	0,384	1	0	
13	H	0,95	0,04	0,109	1	0		0,93	0,07	0,511	0,87	0,12	-0,063	1	0	
14	H	0,88	0,1	0,097	0,94	0,06	0,127	0,87	0,12	0,161	0,73	0,21	0,109	1	0	
15	H	0,92	0,07	0,252	0,94	0,06	0,528*	0,87	0,12	-0,182	0,87	0,12	0,594*	1	0	
16	H	0,87	0,11	0,286*	0,94	0,06	0,455	0,8	0,17	0,373	0,8	0,17	0,337	0,93	0,07	-0,019
17	H	0,92	0,07	0,342**	0,94	0,06	0,528*	0,87	0,12	0,161	0,93	0,07	0,25	0,93	0,07	0,620*
18	H	0,95	0,04	0,051	1	0		0,93	0,07	0,242	0,87	0,12	-0,009	1	0	
23	H	0,44	0,25	0,089	0,63	0,25	-0,223	0,47	0,27	0,214	0,4	0,26	0,422	0,27	0,21	0,064
29	H	0,42	0,25	0,168	0,5	0,27	0,087	0,27	0,21	0,248	0,47	0,27	0,449	0,47	0,27	-0,152
34	H	0,9	0,09	0,099	0,88	0,12	-0,237	1	0		0,87	0,12	0,157	0,87	0,12	0,48
35	H	0,8	0,17	0,350**	0,81	0,16	-0,147	0,8	0,17	0,745**	0,67	0,24	0,462	0,93	0,07	0,262
39	H	0,74	0,19	0,284*	0,69	0,23	-0,179	0,73	0,21	0,429	0,8	0,17	0,617*	0,73	0,21	0,223
46	H	0,61	0,24	0,500**	0,44	0,26	0,424	0,6	0,26	0,38	0,73	0,21	0,555*	0,67	0,24	0,616*
54	H	0,92	0,07	-0,082	0,81	0,16	-0,064	1	0		0,93	0,07	-0,437	0,93	0,07	0,098
55	H	0,92	0,07	0,234	0,88	0,12	0,184	0,93	0,07	0,27	0,93	0,07	0,258	0,93	0,07	0,203

Noot: H= Herkennen; * = Significantie $\leq 0,05$; ** = Significantie $\leq 0,01$; **Zwakke items** = items met variantie $\leq 0,1$ of PB correlatie $\leq 0,2$ in een toetsversie.

Tabel 5.1c Overzicht analyse toetstaak benoemen

BENOEMEN		Algemene Test (N= 61)			FMC (N= 16)			FZC (N= 15)			SMC (N= 15)			SZC (N= 15)		
Itemnr	Type	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr
2	B	0,28	0,2	0,331**	0,25	0,2	0,497	0,13	0,12	0,379	0,4	0,26	0,147	0,33	0,24	0,362
6	B	0,57	0,25	0,582**	0,5	0,27	0,628**	0,6	0,26	0,635*	0,6	0,26	0,552*	0,6	0,26	0,509
11	B	0,66	0,23	0,397**	0,69	0,23	0,446	0,6	0,26	0,649**	0,6	0,26	0,032	0,73	0,21	0,572*
19	B	0,77	0,18	0,349**	0,69	0,23	0,292	0,87	0,12	0,573*	0,8	0,17	0,407	0,73	0,21	0,161
21	B	0,52	0,25	0,308*	0,13	0,12	-0,017	0,33	0,24	0,221	0,87	0,12	0,384	0,8	0,17	0,473
22	B	0,67	0,22	0,375**	0,88	0,12	0,429	0,87	0,12	0,573*	0,53	0,27	0,641*	0,4	0,26	0,349
24	B	0,52	0,25	0,369**	0,56	0,26	0,411	0,33	0,24	0,602*	0,73	0,21	0,407	0,47	0,27	0,097
25	B	0,49	0,25	0,356**	0,56	0,26	0,635**	0,6	0,26	0,399	0,4	0,26	0,498	0,4	0,26	0,002
27	B	0,72	0,2	0,437*	0,56	0,26	0,549*	0,8	0,17	0,540*	0,8	0,17	0,516*	0,73	0,21	0,069
28	B	0,51	0,25	0,163	0,63	0,25	0,44	0,47	0,27	-0,085	0,6	0,26	0,47	0,33	0,24	-0,075
30	B	0,52	0,25	0,448**	0,25	0,2	0,331	0,47	0,27	0,453	0,67	0,24	0,534*	0,73	0,21	0,315
31	B	0,66	0,23	0,219	0,69	0,23	0,274	0,8	0,17	0,239	0,8	0,17	0,24	0,33	0,24	0,353
36	B	0,54	0,25	0,541**	0,5	0,27	0,44	0,53	0,27	0,628*	0,6	0,26	0,590*	0,53	0,27	0,502
37	B	0,56	0,25	0,333	0,56	0,26	0,088	0,87	0,12	0,187	0,47	0,27	0,684**	0,33	0,24	0,607*
38	B	0,61	0,24	0,354**	0,5	0,27	0,348	0,27	0,21	0,419	0,8	0,17	0,229	0,87	0,12	0,442
40	B	0,65	0,23	0,348**	0,56	0,26	0,574*	0,87	0,12	-0,075	0,73	0,21	0,573*	0,47	0,27	0,36
41	B	0,74	0,19	0,492**	0,75	0,2	0,552*	0,73	0,21	0,675**	0,67	0,24	0,344	0,8	0,17	0,463
42	B	0,7	0,21	0,317*	0,69	0,23	0,522*	0,73	0,21	0,328	0,6	0,26	0,256	0,8	0,17	0,178
43	B	0,75	0,18	0,543**	0,69	0,23	0,648**	0,73	0,21	0,326	0,73	0,21	0,617*	0,87	0,12	0,604*
44	B	0,88	0,1	0,303*	0,88	0,12	0,396	0,8	0,17	0,595*	0,87	0,12	-0,019	1	0	
53	B	0,85	0,12	0,464**	0,75	0,2	0,487	0,87	0,12	0,423	0,87	0,12	0,373	0,93	0,07	0,620*

Noot: B = Benoemen; * = Significantie $\leq 0,05$; ** = Significantie $\leq 0,01$; **Zwakke items** = items met variantie $\leq 0,1$ of PB correlatie $\leq 0,2$ in een toetsversie.

Tabel 5.1d *Overzicht analyse toetstaak relaties*

RELATIES		Algemene Test (N= 61)			FMC (N=16)			FZC (N=15)			SMC (N= 15)			SZC (N= 15)		
Itemnr	Type	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr
4	R	0,73	0,08	0,256*	0,65	0,1	-0,133	0,8	0,04	0,259	0,76	0,1	0,294	0,71	0,08	0,643**
5	R	0,36	0,18	0,339**	0,16	0,06	-0,235	0,73	0,17	0,806**	0,4	0,19	0,368	0,17	0,13	0,338
10	R	0,32	0,04	0,600**	0,28	0,02	0,428	0,26	0,03	0,663**	0,42	0,06	0,591*	0,31	0,04	0,749**
20	R	0,16	0,14	0,155	0,19	0,16	0,347	0,2	0,17	0,136	0,07	0,07	0,007	0,2	0,17	0,177
26	R	0,72	0,2	0,280*	0,67	0,06	0,28	0,53	0,09	0,533*	0,67	0,08	0,141	0,62	0,08	0,185
32	R	0,9	0,06	0,339**	0,9	0,06	0,204	0,82	0,11	0,477	0,96	0,03	0,552*	0,91	0,06	0,089
33	R	0,28	0,2	0,24	0,31	0,23	0,377	0,2	0,17	-0,027	0,27	0,21	0,32	0,33	0,24	0,316
45	R	0,42	0,1	0,297*	0,33	0,12	-0,197	0,36	0,09	0,373	0,47	0,09	0,257	0,53	0,11	0,724**
47	R	0,3	0,21	0,657**	0,13	0,12	0,590*	0,4	0,26	0,682**	0,27	0,21	0,719**	0,43	0,25	0,625*
48	R	0,41	0,24	0,509**	0,31	0,23	0,435	0,4	0,26	0,570*	0,47	0,27	0,493	0,47	0,27	0,493
49	R	0,62	0,23	0,527**	0,53	0,25	0,363	0,6	0,26	0,636*	0,73	0,21	0,495	0,63	0,23	0,589*
50	R	0,43	0,21	0,551**	0,39	0,18	0,575*	0,32	0,22	0,599*	0,5	0,25	0,492	0,5	0,21	0,541*
51	R	0,71	0,07	0,398**	0,61	0,08	0,182	0,7	0,07	0,615*	0,78	0,07	0,469	0,77	0,06	0,157
52	R	0,58	0,14	0,427**	0,63	0,18	0,518*	0,47	0,12	0,237	0,6	0,19	0,696**	0,63	0,09	0,155

Noot: R = Relaties; * = Signifcantie niveau $\leq 0,05$; ** = Signifcantie niveau $\leq 0,01$; **Zwakke items** = items met variantie $\leq 0,1$ of PB correlatie $\leq 0,2$ in een toetsversie.

In tabel 5.1a is van de algemene toets en de verschillende toetsversies het testgemiddelde en de variantie weergegeven evenals de waarde van Cronbachs alpha. In de tabellen 5.1b tot en met 5.1d zijn per item het gemiddelde, de variantie en de puntbiseriële correlatie weergegeven.

Uit tabel 5.1a blijkt dat:

1. de toets een hoge alphawaarde heeft van 0,85, hetgeen ruim boven de 0,7 is. De algemene test voldoet daarmee aan de betrouwbaarheidseis.
2. de toetsversies als geheel een hoge alpha hebben, respectievelijk $\alpha_{\text{FMC}} = 0,82$; $\alpha_{\text{FZC}} = 0,88$; $\alpha_{\text{SMC}} = 0,89$; $\alpha_{\text{SZC}} = 0,82$. De alphawaarden voor het onderdeel herkennen (H) voor alle toetsversies laag zijn, respectievelijk $\alpha_{\text{FMC}_H} = 0,14$; $\alpha_{\text{FZC}_H} = 0,63$; $\alpha_{\text{SMC}_H} = 0,64$ en $\alpha_{\text{SZC}_H} = 0,44$. Deze voldoen niet aan de 0,7 ondergrens voor Cronbachs alpha. Dit is een indicatie dat dit toetsonderdeel onvoldoende betrouwbaar is. De geconstateerde alpha-waarden betekenen dat van de toetsversies voor herkennen 14% (FMC), 63% (FZC), 64% (SMC) en 44% (SZC) van de variantie is toe te wijzen aan de variantie van de ware score. Voor alle toetsversies heeft het onderdeel benoemen de hoogste alpha-waarden, respectievelijk $\alpha_{\text{FMC}_B} = 0,81$; $\alpha_{\text{FZC}_B} = 0,80$; $\alpha_{\text{SMC}_B} = 0,82$ en $\alpha_{\text{SZC}_B} = 0,72$ waarbij van elke toetsversie meer dan 70% variantie is toe te wijzen aan de variantie van de ware score. Dit toetsonderdeel is in elke toetsversie betrouwbaar, ook in de algemene toets met een waarde van $\alpha = 0,78$. Voor het onderdeel relaties hebben de toetsversies FZC, SMC en SZC een alphawaarde boven de 0,7, respectievelijk $\alpha_{\text{FZC}_R} = 0,79$; $\alpha_{\text{SMC}_R} = 0,77$ en $\alpha_{\text{SZC}_R} = 0,73$ terwijl de toetsversie FMC een lage waarde ($\alpha_{\text{FMC}_R} = 0,49$) heeft. Daarmee is voor dit toetsonderdeel de toetsversie FMC onvoldoende betrouwbaar in tegenstelling tot de overige versies.

In een variantieanalyse is met de gemiddelde waarden uit tabel 5.1.a nagegaan of er (hoofd)effecten zijn voor de variabelen realisme en context. Uit de ANOVA-analyse blijken er geen hoofdeffecten voor de factoren realisme en context afzonderlijk te zijn en is er ook geen sprake van een interactie-effect tussen de factoren. Dit geldt voor de toets als geheel en voor elke toetstaak. In hoeverre dit resultaat bruikbaar is zal blijken uit de verdere analyse van de toetskwaliteit.

Uit de tabellen 5.1b t/m d zijn hieronder de zwakke items geïnventariseerd:

1. De zwakke items zijn voor herkennen per toetsversie:
  - a. Voor de algemene test de items 3, 7, 8, 9, 12, 13, 14, 15, 17, 18, 23, 29, 34, 54, 55;

- b. voor de toetsversie foto met context (FMC) de items 7, 13, 14, 15, 16, 17, 18, 23, 29, 34, 35, 39, 54 en 55;
  - c. voor de toetsversie foto zonder context (FZC) de items 1, 7, 8, 9, 12, 13, 14, 15, 17, 18, 34, 54 en 55;
  - d. voor de toetsversie schema met context (SMC) de items 3, 9, 13, 14, 17, 18, 34, 54 en 55;
  - e. voor de toetsversie schema zonder context (SZC) de items 3, 7, 9, 12, 13, 14, 15, 16, 17, 18, 23, 29, 35, 54 en 55.
2. De zwakke items voor benoemen zijn per toetsversie :
- a. voor de algemene test de items 28 en 44;
  - b. voor de toetsversie foto met context (FMC) de items 21 en 37;
  - c. voor de toetsversie foto zonder context (FZC) de items 28, 37 en 40;
  - d. voor de toetsversie schema met context (SMC) de items 2, 11 en 44;
  - e. voor de toetsversie schema zonder context (SZC) de items de items 19, 24, 25, 27, 28, 42, 44 en 53.
3. De zwakke items voor relaties zijn per toetsversie:
- a. voor de algemene test de items 4, 10, 20, 32, 45 en 51;
  - b. voor de toetsversie foto met context (FMC) de items 4, 5, 10, 26, 32, 45 en 51;
  - c. voor de toetsversie foto zonder context (FZC) de items 4, 10, 20, 26, 33, 45 en 51;
  - d. voor de toetsversie schema met context (SMC) de items 4, 10, 20, 26, 32, 45 en 51;
  - e. voor de toetsversie schema zonder context (SZC) de items 4, 10, 20, 26, 32, 51 en 52.

Om vanuit een andere invalshoek de toets met bijbehorende items te onderzoeken is de Mokkenanalyse uitgevoerd die staat weergegeven in tabellen 5.2 a & b. Deze analyse moet inzicht verschaffen in hoeverre items gezamenlijk hetzelfde construct meten. Deze analyse is op alle items uitgevoerd zonder differentiatie naar de toetsversies. Daarmee is ervoor gekozen om de schaalbaarheid van de items te onderzoeken over de gehele populatie. De steekproefgrootte was dusdanig dat het niet opportuun was om de analyse ook nog uit te voeren voor de verschillende toetsversies.

Tabel 5.2a Resultaten Mokkenanalyse op toets en deoltoetsen

	Schaal 1	Schaal 2	Schaal 3	Schaal 4	Schaal 5	Niet schaalbaar
Gehele toets	34, 35	1, 3, 4, 9, 21, 24, 25, 26, 40, 49, 51, 53, 54,	28, 29	17, 39	38, 42	2, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 22, 23, 27, 30, 31, 32, 33, 36, 37, 41, 43, 44, 45, 46, 47, 48, 50, 52, 55
Herkennen	46, 54, 55					1, 3, 7, 8, 9, 12, 13, 14, 15, 16, 17, 18, 23, 29, 34, 35, 39
Benoemen	25, 53	11, 19, 21	38, 40, 42	36, 37	43, 44	2, 6, 22, 24, 27, 28, 30, 31, 41
Relaties	33, 49, 50, 51, 52	10, 48				4, 5, 20, 26, 32, 45, 47,

Tabel 5.2b Overeenkomsten en verschillen geschaalde items op toetstaak en instructiethema

Gehele toets	Geschaalde items	
Schaal 1	34, 35	Dezelfde toetstaak en hetzelfde instructiethema
Schaal 2	1, 3, 4, 9, 21, 24, 25, 26, 40, 49, 51, 53, 54,	Verschillende toetstaken en verschillende instructiethema's (toetstaak: (h) 1 & 3 & 9 & 54, (b) 21 & 24 & 25 & 40 & 53, (r) 4 & 26 & 49 & 51) (thema: buikholte & spijsvertering 1, 3, 4 & 9, geslachtsorganen 21, 24, 25 & 26, borstholte 40 & 49, (bij)nieren & bloedvaten 51, 53 & 54)
Schaal 3	28, 29	Verschillende toetstaak (benoemen & herkennen) en hetzelfde instructiethema (geslachtsorganen)
Schaal 4	17, 39	Dezelfde toetstaak (herkennen) en verschillende instructiethema's (spieren & borstholte)
Schaal 5	38, 42	Dezelfde toetstaak (benoemen) en hetzelfde instructiethema (borstholte)
<i>Herkennen</i>		
Schaal 1	46, 54, 55	Dezelfde toetstaak (herkennen) en twee instructiethema's (borstholte 46 en (bij)nieren & bloedvaten 54 & 55)
<i>Benoemen</i>		
Schaal 1	25, 53	Dezelfde toetstaak en verschillende instructiethema's (geslachtsorganen en (bij)nieren en bloedvaten)
Schaal 2	11, 19, 21	Dezelfde toetstaak en hetzelfde instructiethema (geslachtsorganen)
Schaal 3	38, 40, 42	Dezelfde toetstaak en hetzelfde instructiethema (borstholte)
Schaal 4	36, 37	Dezelfde toetstaak en hetzelfde instructiethema (borstholte)

Tabel 5.2b *Overeenkomsten en verschillen geschaalde items op toetstaak en instructiethema*

Gehele toets	Geschaalde items	
Schaal 5	43, 44	Dezelfde toetstaak en hetzelfde instructiethema (borstholte)
<i>Relaties</i>		
Schaal 1	33, 49, 50, 51, 52	Dezelfde toetstaak en drie verschillende instructiethema's (geslachtsorganen 33, borstholte 49 & 50, (bij)nieren en bloedvaten 51 & 52)
Schaal 2	10, 4	Dezelfde toetstaak en hetzelfde instructiethema (buikholte en spijsvertering)

De schalen in tabel 5.2a zijn samengesteld op basis van een sterke samenhang tussen de items die hetzelfde construct zouden meten en tabel 5.2b geeft aan of de gevonden schaal herleidbaar is naar een gemeenschappelijke toetstaak en/of een instructiethema met het volgende resultaat:

1. Voor de gehele toets zijn er items schaalbaar in vijf schalen. Daarvan zijn vier schalen van elk twee items met dezelfde toetstaak en/of hetzelfde instructiethema.
2. Voor de toetstaak 'herkennen' zijn drie items schaalbaar (46, 54 en 55). Eén item (54) bleek in de eerdere analyses al een zwak item te zijn. De andere twee items behoorden niet tot hetzelfde instructiethema. Alle andere items bleken niet schaalbaar en dat bevestigt dat de toetstaak herkennen nog onvoldoende van kwaliteit is en bijstelling vereist.
3. Voor de toetstaak benoemen zijn er vijf schalen, te weten schaal 1: items 25 en 53; schaal 2: items 11, 19, en 21; schaal 3: items 38, 40 en 42; schaal 4: items 36 en 37 alsook schaal 5: items 43 en 44. Uit eerdere analyse bleek 44 echter een zwak item voor de toetsversie SZC. Dat de items die in eenzelfde schaal zitten mogelijk hetzelfde construct meten is afleidbaar voor de schalen 3, 4 en 5 waarin de items hetzelfde instructiethema toetsen. Dit resultaat ondersteunt wellicht de eerdere resultaten waarin de toetstaak 'benoemen' voor elke toetsversie een alphawaarde heeft die voldoet aan de norm, er geen opvallende zwakke of sterke items zijn en de items onderscheidend blijken te zijn.
4. Voor de toetstaak 'relaties' zijn er twee schalen, te weten schaal 1: items 33, 49, 50, 51 en 52 en schaal 2: items 10 en 48. Opvallend is item 51 dat in eerdere analyses een zwak item bleek voor de toetsversies FMC en SZC.

De analyses leveren tot nog toe geen resultaten die wijzen op een kwalitatief goede toets die inhoudelijk goed gedekt is en daarmee het beoogde construct meet. Ondanks dat de alphawaarden voor de afzonderlijke toetsen voldoende zijn, blijken de toetsen veel zwakke items te bevatten. Geen van de toetsversies met de toetstaak herkennen voldoen aan de minimale ondergrens voor Cronbachs alpha evenals de alphawaarde voor de toetstaak relaties in de FMC toetsversie. Een meerderheid aan items blijkt niet schaalbaar. Dit resultaat noodzaakt tot aanpassing van de toets, met name in de onderdelen herkennen en relaties.

Voor wat betreft de vergelijkbaarheid van de toetsversies is nagegaan of de correlaties tussen de toetstaken/ (deel)constructen per toetsversie vergelijkbaar zijn. Dit leverde vier correlatiematrices op die in tabel 5.3 in zijn weergegeven. Het betreft de correlatie van de totaal-scores op de drie toetstaken voor iedere toetsversie. De getallen met een *- of een **-aanduiding geven aan dat de correlatie tussen twee toetstaken significant is, met significantieniveaus van respectievelijk ($p \leq 0,05$) en ($p \leq 0,01$).

Tabel 5.3 *Correlatiematrices van totaal scores op Herkennen, Benoemen en Relaties*

FMC	Herkennen	Benoemen	Relaties
Herkennen	1,000	0,577*	0,289
Benoemen	0,577*	1,000	0,603
Relaties	0,289	0,603*	1,000
FZC	Herkennen	Benoemen	Relaties
Herkennen	1,000	0,521*	0,430
Benoemen	0,521*	1,000	0,735**
Relaties	0,430	0,735**	1,000
SMC	Herkennen	Benoemen	Relaties
Herkennen	1,000	0,672**	0,687**
Benoemen	0,672**	1,000	0,524*
Relaties	0,687**	0,524*	1,000
SZC	Herkennen	Benoemen	Relaties
Herkennen	1,000	0,381	0,542*
Benoemen	0,381	1,000	0,516*
Relaties	0,542*	0,516*	1,000

Noot: * = $p < 0,05$; ** = $p < 0,01$; FMC = testvorm foto's met context; SMC = testvorm schematisch met context; FZC = testvorm foto's zonder context; SZC = testvorm schematisch zonder context, H = herkenningstaak; R = functionele relatietaak; B = benoemingstaak.

Uit tabel 5.3 blijkt dat er binnen elke toetsversie significante correlaties zijn tussen de toetstaken benoemen en relaties. Bij FMC, SMC en FZC zijn er significante correlaties tussen de taken herkennen en benoemen. Bij SMC en SZC zijn er significante correlaties tussen de taken herkennen en relaties.

Met de vier correlatiematrices is vervolgens nagegaan in hoeverre deze met elkaar overeenkomen. Daartoe zijn, onder de assumptie dat de scores multivariaat normaal verdeeld zijn, modellen voor de vier correlatiematrices opgesteld waarvoor de log-likelihood en de bijbehorende deviance zijn berekend. Het eerste model beantwoordde aan de hypothese dat alle vier matrices gelijk zijn. Voor het laatste model gold de hypothese dat alle vier matrices verschillend zijn. In totaal zijn 15 modellen mogelijk waarvan in tabel 5.4 de vier meest waarschijnlijke staan weergegeven. De log-likelihood is de kans dat het model dekkend is voor de geobserveerde data en wordt weergegeven met de deviance. Deze laatste heeft als waarde min twee maal de log-likelihood. Hoe hoger de log-likelihood, respectievelijk hoe lager de deviance, hoe waarschijnlijker en acceptabeler het model over de vergelijkbaarheid van de correlatiematrices. Het best passende model heeft dus de kleinste deviance en dat betekent de grootste waarschijnlijkheid dat de hypothese over de vergelijkbaarheid van de correlatiematrices bij de data past. In tabel 5.4 zijn de vier modellen weergegeven met de kleinste waarden voor de deviance. Daarmee zijn ze tevens de vier meest waarschijnlijke modellen.

Tabel 5.4 *Onderzoek gelijkheid correlatiematrices*

FMC	FZC	SMC	SZC	Deviance
3	1	2	1	833.14
3	3	2	1	834.51
1	3	2	1	835.50
3	2	1	1	837.04

Noot: Gelijke correlatiematrices hebben hetzelfde nummer in de eerste vier kolommen.

In tabel 5.4 geldt voor het model op de eerste rij de hypothese dat de matrices FZC en SZC aan elkaar gelijk zijn, en dat FMC en SMC daarvan en van elkaar verschillen. Daarna volgt op de tweede rij het model waarbij FMC en FZC dezelfde matrices hebben en de rest unieke matrices. In de derde rij volgt het model dat FMC en SZC gelijk zijn en de overige matrices uniek en tot slot het model in de vierde rij dat de matrices SMC en SZC aan elkaar gelijk zijn en de andere verschillen. Dit levert tot eindresultaat dat zowel de hypothese dat alle vier de matrices verschillen als de hypothese dat ze alle vier hetzelfde waren, moet worden verworpen. De hypothese dat die structuur voor alle vier de

toetsversies hetzelfde was, kon danook niet worden bevestigd evenals de hypothese dat de structuur voor alle vier de toetsversies verschilde. Overigens dient te worden opgemerkt dat de hier onderzochte geobserveerde correlaties verlaagd zijn (geattenuëerd) door de onbetrouwbaarheid van de toetsen. In het volgende hoofdstuk wordt de correlatiestructuur verder onderzocht met vier aangepaste toetsen, een grotere steekproef en een gedetailleerder statistisch model, namelijk een parametrisch IRT-model waarbij gecorrigeerd wordt voor de onbetrouwbaarheid van de toetsen.

5.5 DISCUSSIE EN CONCLUSIES

Het in dit hoofdstuk gepresenteerde exploratieve experiment had tot doel de kwaliteit van de eerste versie van de CBT-toets te onderzoeken die in het onderzoek is ontwikkeld als onderdeel van de instructie-toetscombinatie waarmee het in dit project aangepakte toetsprobleem moet worden opgelost. Uit de resultaten blijkt dat er duidelijk aanleiding is tot verbetering alsook verdere ontwikkeling van de toets. Vooral het toetsonderdeel 'herkennen' blijkt zwak en dat geldt voor alle vier toetsversies. Het toetsonderdeel 'benoemen' blijkt in alle vier toetsversies het meest stabiele onderdeel. Het toetsonderdeel 'relaties' lijkt voldoende behalve voor de toetsversie foto met context. Verder blijkt dat de toetsversie schema zonder context de meeste zwakke items bevat.

Het schalen van de items met de Mokkenanalyse is op een kleine steekproef gebaseerd, dus conclusies moeten met voorzichtigheid worden getrokken. Maar weinig items lijken samen te hangen. Het toetsonderdeel 'benoemen' levert de meeste schaalbare items die net als de schaalbare items bij de andere onderdelen verdeeld zijn over meerdere schalen. Voor de gevonden schalen lijkt vooralsnog het gemeenschappelijk hebben van de toetstaak en/of het instructiethema de samenhang te bepalen. Het aantal items in de schalen is te gering om een trend te ontdekken die leidt tot een duidelijk beeld van het construct of het samenstel van deelconstructen dat de verworven kennis representeert.

Samenvattend zijn de resultaten niet erg eenduidig om uitspraken te doen over de mate waarin de items gezamenlijk dekkend zijn voor het adequaat toetsen van de visuele (deel)constructen. Uitspraken over welke representatievorm de (deel)constructen het beste toetst is niet mogelijk, waarbij wordt opgemerkt dat de

toetsversie schema met context de voorkeur heeft. Studenten scoren het best op deze toetsversie die zowel op de toets als op de (deel)toetsen hogere Cronbachs alphaswaarden heeft in vergelijking tot de toetsversies foto met context en schema zonder context. Voor wat betreft de representatievormen blijken er geen hoofd- en interactie-effecten te zijn voor de factoren realisme en context.

Met de studentassistenten is het verloop van de toetsafname mondeling geëvalueerd waarbij is ingegaan op het doorlopen protocol ter voorbereiding van de studenten op de toets. Deze merkten op dat vanuit de zaal door studenten is aangegeven dat ondanks de mondelinge bespreking van oefenitems, vooraf aan de toets, er nog onduidelijkheden waren over de wijze van invulling van de items in de toets. Hoewel dit niet expliciet gemeten is, is dit een punt van zorg. In het volgende experiment dient de procedure op dit onderdeel te worden verbeterd.

De eindconclusie is dat de toets zoals toegepast in het hier gepresenteerde experiment nog van onvoldoende kwaliteit is en dient te worden bijgesteld. Daarvoor is evaluatie van de toets noodzakelijk om tot gefundeerde bijstellingen te komen. Daarna kan met de bijgestelde toets het experiment nogmaals worden uitgevoerd. In de volgende paragraaf wordt het hier gepresenteerde hoofdstuk afgerond met de evaluatie en bijstelling van de toets naar aanleiding van de resultaten uit het experiment.

5.6 BIJSTELLING VAN DE TOETS

Deze evaluatie van de toets is gericht op de inhoudelijke dekking van het kennisgebied, de wijze waarop de visuele (deel)constructen zijn geoperationaliseerd naar de toetstaken (herkennen, benoemen en relaties) evenals de operationalisatie van de factoren (realisme en context). Daartoe worden:

1. vooral de zwakke toetsitems uit de toets geëvalueerd en zo nodig bijgesteld.
2. voor de te toetsen visuele (deel-)constructen nagegaan in hoeverre elk construct gedekt is door het overeenkomstig toetsonderdeel;
3. voor de factoren realisme (foto's of schema's) en context (al-dan-niet) nagegaan in hoeverre deze voldoende zijn verwerkt in de toets.

Aanpassingen in de operationalisatie van de variabelen en de toetstaken

Voor de factor realisme bleek uit het experiment dat er geen significant verschil was tussen foto en schema voor de toets als geheel en voor elke toetstaak. Na bestudering van de operationalisatie van de factor realisme is besloten geen veranderingen aan te brengen in de bestaande representatievormen omdat de onderzoeker en de vakdocent het gezamenlijk eens waren dat de foto's het meest de onderwijsleersituatie benaderen en de schema's met reële objectcontouren voldoende informatie verschaffen over belangrijke objectkenmerken. Verdere vereenvoudiging van schematische representatievormen vergroot de kans dat de aangepaste schema's te abstract zullen zijn. De gemeenschappelijke conclusie is dat de voor de in de CBT toets gebruikte plaatjes beide representatievormen al uiterste vormen zijn van de factor realisme namelijk van een heel gedetailleerde representatievorm (de foto's) tot een heel schematische representatievorm (de contourtekeningen). Daarbij wordt vastgehouden aan een abstractieniveau van schematische weergaven waarin anatomische contouren duidelijk herkenbaar zijn. Bijkomend argument is dat de toetsversie schema zonder context de meest zwakke toets bleek en mede daardoor is het niet verantwoord deze vormen aan te passen door de schematische vorm nog abstracter weer te geven.

Voor de factor context zijn wel veranderingen aangebracht. Uit de resultaten bleken er geen significante verschillen te zijn tussen toetsversies met en de toetsversies zonder context. Bestudering van de toetsversies met de vakdocent leidde tot de conclusie dat deze factor mogelijk onvoldoende is vorm gegeven omdat in de toetsvragen alle anatomische structuren feitelijk steeds zonder context zijn weergegeven. Bij de operationalisatie van de factor context in de toetsversies is de contextuele informatie toegevoegd door het plaatsen van een apart kader met daarin de contouren van de rat waarin voor elk toetsitem de exacte positie van de betreffende anatomische structuur is weergegeven. Doordat er tijdens het experiment geen verschillen zijn aangetoond tussen de aan- of afwezigheid van context is bestudeerd op welke wijze deze representatievormen nog meer van elkaar kunnen worden onderscheiden. De toetsversie met context is voor elke toetstaak aangepast zodat deze zich meer onderscheidt van de toetsversie zonder context:

1. Voor de toetstaak herkennen is getracht elk antwoordalternatief te visualiseren binnen de contouren van de rat, in relatie tot daar omheen liggende objecten. In geval het niet mogelijk is elk antwoordalternatief te tonen binnen de context, wordt de context in een apart kader weergegeven waarbij ook de omliggende objecten binnen de contouren van de rat worden getoond.

2. Voor de toetstaak benoemen is getracht het gevraagde object te tonen binnen de contouren van de rat, inclusief alle objecten die binnen deze contouren liggen, zoals geobserveerd tijdens de dissectie.
3. Tot slot is voor de toetstaak relaties de context steeds weergegeven als een apart kader met de contouren van de rat en de objecten binnen die contouren zoals geobserveerd tijdens de dissectie.

Aanpassing van de representatievorm voor elke toetstaak heeft als voordeel dat de verschillen tussen de toetsversies met en zonder contextuele informatie maximaal worden vormgegeven. Figuur 5.3 geeft deze veranderingen weer. De foto's in de tweede rij laten de oorspronkelijke representatievorm met context zien. De foto's in de derde rij tonen de aangepaste representatievorm waarin de structuur zoveel mogelijk in de context, de contouren van de rat, is geplaatst zoals geobserveerd tijdens de dissectie. In de aangepaste representatievorm met context hebben de studenten, bij benadering, alle cues zoals zij die waarnamen tijdens de dissectie van de rat.

	Herkenningsstaak	Benoemingsstaak	Relatietaak
<i>Oorspronkelijke toetsversie</i>			
<i>Aangepaste toetsversie</i>			

Figuur 5.3 Visualisatie van de conditie 'met context' in de oorspronkelijke en in de aangepaste toetsversie

Figuur 5.4 Overzicht van de verschillende representatievormen van de te toetsen constructen

Deze wijziging in de operationalisatie van de factor context heeft tot gevolg dat de toetsversies met context, FMC en SMC, geheel zijn gewijzigd voor alle items. Figuur 5.4 biedt een overzicht van de toetsversies zoals die nu zijn vormgegeven voor elke toetstaak.

Vergelijking van figuur 5.4 met figuur 4.5 geeft een precies beeld van de aangepaste representatievormen in vergelijking tot de oorspronkelijke representatievormen.

Dekking (deel-)construct in de toets

Na het bijstellen van de representatievormen is de toets inhoudelijk bestudeerd en aangepast in samenspraak met de vakdocent. Met de vakdocent is de volgende werkwijze gehanteerd:

1. Items die naar aanleiding van de experimentele resultaten als zwak te boek staan zijn bestudeerd op
  - a. de vorm van de vraagstelling alsook daarbij horende antwoordalternatieven.
  - b. de bijdrage van de items aan de betreffende toetstaak.
  - c. de mate waarin dit item, in combinatie met gerelateerde items, dekkend is voor een specifiek dissectiethema.
2. De toetstaken zijn afzonderlijk bestudeerd waarbij is gelet op
  - a. de bijdrage aan de toets gebruikmakend van de analyse van tabel 5.1a tot en met tabel 5.1.d.
  - b. de redactie van de vraagstelling en antwoordalternatieven van items binnen de toetstaak.
  - c. de foutieve antwoorden die zijn gegeven in het experiment.

Deze werkwijze leverde de volgende resultaten op:

Voor de toetstaak herkennen bleek voor elke toetsversie de betrouwbaarheid te laag ($\leq 0,7$) hetgeen werd veroorzaakt door het grote aantal zwakke items. Deze items zijn bestudeerd en zes zijn aangepast op (a) de visualisatie van de variabele context zoals beschreven in de vorige paragraaf, (b) de vraagvorm, deze werd redactioneel verbeterd, en (c) de antwoordalternatieven als er sprake was construct-irrelevante gemakkelijkerheid. Item 18 is verwijderd uit de toets. De overige items zijn, na overleg met de vakdocent, niet gewijzigd. Uit overleg met de vakdocent bleken deze items relevant voor het dekken van het inhoudsgebied waarbij optionele aanpassingen niet resulteerden in kwaliteitsverbetering. Er bleven 19 items over waarna is gekeken of deze tezamen dekkend zijn voor het kennisgebied. Dat bleek het geval met als resultaat dat het onderdeel herkennen in de aangepaste toets 19 items bevat.

Voor de toetstaak benoemen zaten in de eerste CBT-toetsversie zowel foutief als correct geschreven Latijnse benamingen. Na bestudering van de items en de gemaakte fouten werd duidelijk dat er met de benoemingstaak twee verschillende eigenschappen werden getoetst namelijk (a) de mate waarin een kandidaat de correcte Latijnse benaming van een object kent en (b) de correcte schrijfwijze kent. Hiervoor is eerder gekozen naar aanleiding van de oude toetsmethode waarin de benamingen van anatomische onderdelen in tekeningen werden beoordeeld op zowel correcte spelling als benaming. Uit een analyse van de fouten die kandidaten maakten bij deze taak, bleek dat zij veelal items fout benoemden door de verkeerde benaming in te vullen en niet door een correcte maar foutief

geschreven benaming. Niet alleen bleek spelling geen probleem maar toetstechnisch is het niet terecht twee eigenschappen in hetzelfde item te toetsen. In samenspraak met de vakdocent is ervoor gekozen om maar één eigenschap te toetsen namelijk of de kandidaat al dan niet de correcte Latijnse benaming van een object kent. De toets is voor deze taak aangepast door alleen correct geschreven Latijnse benamingen in de index te plaatsen. Voorts zijn drie items (24, 25 & 33) redactioneel of in visualisatie gewijzigd en is item 44 uit de toets verwijderd. Voor de aangepaste toets is het eindresultaat dat het onderdeel benoemen 20 items bevat.

De toetstaak relaties bleek betrouwbaar te zijn voor alle toetsversies met uitzondering van de toetsversie foto met context. De gemiddelde scores op de items zijn lager dan die op items van de andere toetstaken. Een mogelijke oorzaak is dat de items ‘moeilijker’ zijn dan die in de andere toetstaken. Een andere mogelijke oorzaak kan liggen in de wijze van scoring. In dit experiment was het bij de relatietaken mogelijk om in een item meerdere objecten met elkaar te verbinden. Daarbij verschilde het aantal gevraagde verbindingen per item waarmee in de scoring weinig rekening werd gehouden. Een foutieve verbinding betekende een fout item. Uit de resultaten bleken meerdere studenten op enkele items waarin meerdere verbindingen werden bevraagd slecht te presteren wat het beeld gaf van constructirrelevante moeilijkheid. Deze items zouden onterecht moeilijk zijn vanwege de scoringsmethodiek. Om deze reden is besloten voor deze taak in elk item maar naar één enkele verbinding of twee verbindingen te vragen zodat de kandidaat maximaal twee handelingen hoeft uit te voeren. Dit is niet alleen terecht voor wat betreft de vraagstelling maar heeft een bijkomend positief effect namelijk dat dit type item ook beter scorbaar is. In tegenstelling tot hoofdstuk 5 zijn voor items met een relatietaak alle items dichotoom gescoord. Als gevolg hiervan zijn de items 5 en 45 verwijderd, drie items (20, 51, 52) redactioneel gewijzigd en zes nieuwe items geconstrueerd. Het eindresultaat is een aangepaste toets met 18 items voor de toetstaak functionele relaties.

Tot slot is een eindcontrole uitgevoerd met als doel na te gaan of de toets na de wijzigingen nog dekkend is voor het gehele geleerde construct op basis van de vijf instructiethema's, zoals besproken in hoofdstuk 4, en de te toetsen deelconstructen, gerelateerd aan de toetstaken.

In onderstaand overzicht, figuur 5.5, is van de bijgestelde toets de dekking weergegeven naar aanleiding van de invulling van de toets.

	Toetstaken	Herkennen	Benoemen	Relaties	Per thema
Themanr	<i>Instructiethema's</i>				<i>Aantal items</i>
1	Buikholte & spijsvertering	1, 3, 9, 10, 14	2, 8, 13	4, 5, 6, 7, 11, 12	14
2	Ventrale spieren	15, 16, 17, 18, 19, 20			6
3	Geslachtsorganen	25, 31	21, 23, 24, 26, 27, 29, 30, 32, 33	22, 34, 35	14
4	Borstholte en grote bloedvaten	36, 37, 41, 49	38, 39, 40, 42, 43, 44, 45	46, 47, 48, 50, 51, 52, 53	18
5	(bij)nieren en bloedvaten	57, 58	56	54, 55	5
Totaal		19	20	18	57

Figuur 5.5 Overzicht items in de bijgestelde toets in relatie tot dekking van het construct

Uit bovenstaand figuur blijkt dat de toets weliswaar geheel dekkend is voor de thema's die aan bod komen maar dat voor het thema ventrale spieren niet wordt getoetst op het kunnen benoemen van de spieren en kunnen leggen van functionele relaties tussen de spieren onderling. Met de docenten is inhoudelijk overlegd over dit thema, het te toetsten construct en de daaruit voortvloeiende toetsingsmogelijkheden. Resultierend uit de inhoudelijke analyse achtten docenten de herkenning van de spieren voldoende voor de toetsing van het construct.

Samengevat zijn er voor de toets drie integrale aanpassingen, te weten: (a) de toetsversies met context (FMC en SMC) onderscheiden zich nog meer van de toetsversies zonder context (FZC en SZC) doordat de visualisaties met context gedetailleerder zijn gevisualiseerd, (b) de benoemingstaak is ingeperkt door in de antwoordalternatieven enkel correct geschreven Latijnse benamingen op te nemen en (c) de relatietask is ingeperkt tot het leggen van één à twee verbindingen tussen objecten. Daarnaast zijn er gedetailleerdere aanpassingen waarbij twaalf items redactioneel zijn gewijzigd, zes nieuwe items zijn geconstrueerd en enkele items zijn verwijderd. Het eindresultaat is een aangepaste toets van 57 items (55 in de vorige versie), waarvan voor herkenning 19 items (20 in de vorige versie), voor benoemen 20 items (21 in de vorige versie) en voor relaties 18 items (14 in de vorige versie). Hierbij dient te worden opgemerkt dat de toets weliswaar bestaat uit 57 items maar in de nummering is

ook een item 58 aanwezig en ontbreekt item 28 omdat deze laatste gedurende het ontwikkelproces is verwijderd en niet vervangen.

Het eindresultaat is een aangepaste toets die compleet is weergegeven in de bijlage bij hoofdstuk 5. Deze toets is in deze vorm niet meer één op één vergelijkbaar met de eerdere versie uit hoofdstuk 4. Om een beeld te krijgen van de verschillen tussen de aangepaste toets en de eerdere versie moeten de bijlagen van hoofdstuk 5 en hoofdstuk 4 met elkaar worden vergeleken. Met deze aangepaste toets is het experiment zoals beschreven in hoofdstuk 5 nogmaals uitgevoerd, waarover wordt gerapporteerd in hoofdstuk 6.

HOOFDSTUK 6

Visuele CBT-toets – 2e experiment

In dit hoofdstuk wordt de toets uit hoofdstuk 4 die naar aanleiding van de resultaten uit hoofdstuk 5 op een aantal punten is bijgesteld, opnieuw onderzocht. Daartoe wordt het experiment van hoofdstuk 5 nog eens uitgevoerd met als hoofddoel het evalueren van de ontwikkelde instructie-toetscombinatie, de eerste onderzoekslijn. Langs de tweede onderzoekslijn wordt onderzocht welke vormgeving van de visuele representaties in termen van realisme (schema's of foto's) en context (wel of geen) tot de meest betrouwbare toets leidt, en of de verschillende versies van de toets dezelfde constructen Herkennen, Benoemen en Relaties meten.

6.1 INTRODUCTIE

De psychometrische analyses ter beoordeling van de kwaliteit van de toets worden in dit experiment net zo uitgevoerd als in het vorige experiment. Er vindt een toets en item analyse (TIA) plaats op basis van klassieke testtheorie en de verschillen in de moeilijkheidsgraad en de betrouwbaarheid van de verschillende toetsversies worden vergeleken met behulp van variantieanalyse (ANOVA, MANOVA). De invloed van de factoren realisme en context wordt met behulp van ANOVA en MANOVA geanalyseerd. De vraag of de verschillende toetsversies vergelijkbare constructen meten wordt onderzocht met behulp van item response theorie (IRT). Tot slot vindt er een protocolanalyse plaats naar de kwaliteit van de instructie-toetscombinatie vanuit studentperspectief.

6.2 HET EXPERIMENT

Onderzoeksdesign

In het experiment wordt de bijgestelde toets gebruikt in de vier combinaties van de variabelen realisme en context die de genoemde vier experimentele condities

vormen. Het experiment vindt plaats volgens het 2*2 factoriële ontwerp dat is weergegeven in figuur 6.1.

		Context	
		Wel	Geen
Realisme	Foto	FMC	FZC
	Schema	SMC	SZC

Figuur 6.1 Overzicht variabelen realisme en context met de combinatie in visualiteitsvormen

Het onderzoeksdesign, weergegeven in figuur 6.2, bevat vier experimentele condities. Deze condities komen overeen met de combinaties in representatievormen die eerder in figuur 6.1 staan weergegeven. De experimentele condities zijn: kleurenfoto's met context (FMC), kleurenfoto's zonder context (FZC), schema's met context (SMC) en schema's zonder context (SZC).

Voor dit experiment is gekozen om respondenten te vragen op welke wijze zij een item hebben beantwoord in de toets door ze na het invullen van de toets hun ingevulde item nogmaals te tonen met de vraag hoe zij tot het antwoord zijn gekomen. Het doel is kwalitatieve informatie te vergaren over de instructie-toetscombinatie waarbij de onderliggende processen bij het beantwoorden van een toetsitem, in relatie tot de instructie, in kaart worden gebracht. Hierbij wordt nagegaan naar welke (onderdelen van de) instructie respondenten verwijzen bij het oplossen van een toetsitem. De gekozen methodiek is de protocolanalyse (PA) (Messick, 1993) waarin aan respondenten wordt gevraagd om retrospectief alle stappen te beschrijven die ze hebben genomen om tot een geschikte oplossing te komen. In de beschrijving van de gevolgde procedure wordt hier dieper op ingegaan.

R	X	O	FMC	PA
	X	O	FZC	PA
	X	O	SMC	PA
	X	O	SZC	PA

Noot: at random toewijzing van studenten in de groepen; X: Instructie "Anatomie van de Rat"; O: Oefening toetsprogramma; FMC: Afname toetsversie kleurenfoto met context; FZC: Afname toetsversie kleurenfoto zonder context; SMC: Afname toetsversie schema met context; SZC: Afname toetsversie schema zonder context; PA: Protocol analyse.

Figuur 6.2 Onderzoeksdesign

Procedure

Het experiment is uitgevoerd in het eerste curriculumjaar van de opleiding biologie bij de afdeling dierfysiologie. De procedure rondom het experiment is grotendeels gelijk aan de procedure van het experiment zoals eerder beschreven in hoofdstuk 5.

De onderzoeksgroep bestond uit 134 eerstejaars biologiestudenten. Deze studenten hadden enige dissectievaardigheden opgedaan met een haai en met een duif maar niet met de rat. Vooraf aan de instructie werden studenten eerst at random toegewezen aan een practicumgroep en aan een van de vier experimentele condities.

In overeenstemming met de procedure zoals aangegeven in hoofdstuk 5 kregen studenten voor de instructie een introductie met informatie over de dissectie-instrumenten, de practicumprocedure en het veiligheidsreglement. De studenten hadden geen voorkennis van de experimentele condities. Voor alle groepen duurde de instructie twee dagen van elk acht uur. Per dag werd één van de instructiehelften doorlopen en op de tweede dag werd de toets afgenomen over de instructieonderwerpen van de twee dagen. Bij dit experiment is besloten te toetsen aan het eind van de tweede instructiedag omdat verwacht wordt dat de kennis van de eerste instructiedag nog voldoende recent is om op de tweede dag te worden getoetst. Dit is mede ingegeven door het feit dat er in de experimenten van de hoofdstukken 3 en 5, waarin de deelttoets over de eerste instructiehelft op de eerste dag werd afgenomen en de deelttoets over de tweede instructiehelft op de tweede dag, geen significante verschillen zijn gevonden tussen de toetsprestaties op de twee deelttoetsen, waardoor de resultaten van de twee dagen konden worden samengevoegd.

Voor het volgen van de instructie werden de studenten ingedeeld in groepjes van 10 à 12 personen. Elke groep kreeg een studentassistent toegewezen als docent die de instructie gaf. In totaal waren er zes studentassistenten die de instructie gaven. Deze assistenten waren door de onderzoeker getraind en hadden geen voorkennis over de inhoud van de toetsen. Het practicum is tweemaal uitgevoerd in twee groepen van elk 67 studenten.

Er zijn wel enkele aanpassingen verricht in de procedure in vergelijking tot het eerste experiment.

Allereerst bleek tijdens het experiment van hoofdstuk 5 dat ondanks de mondelinge bespreking van oefenitems, vooraf aan de toets, er nog onduidelijkheden waren over de wijze van invulling van deze items. Door de studenten dit keer daadwerkelijk te laten oefenen met oefenitems is geprobeerd die onduidelijkheden op te heffen. Zij kregen van elke toetstaak een voorbeelditem dat zij moesten oplossen. Voorts moesten de studenten na het invullen van de toets het computerprogramma vervolgen waarbij zij een aantal items die zij hadden beantwoord nogmaals kregen voorgelegd ten behoeve van een protocolanalyse. De reden voor de keuze van een retrospectieve benadering is ter voorkoming dat deze evaluatiemethode zou kunnen interfereren met de wijze waarop de toets werd gemaakt. Studenten konden schriftelijk of mondeling reageren. In de schriftelijke variant kregen 88 studenten na het invullen van de toets de helft van de door hen ingevulde toetsvragen met de opdracht per item in te typen op welke wijze zij deze hebben beantwoord. De selectie van toetsitems bestond uit de even genummerde items of de oneven genummerde items. Bij de randomisatie van de studenten is rekening gehouden met de intentie elf antwoorden voor elk item in elke conditie te vergaren. De resterende groep van 46 studenten kreeg de mondelinge variant waarbij met een hoofdtelefoon met microfoon de vragen bij de items mondeling werden gepresenteerd en zij hun antwoord konden inspreken. Verwacht werd dat het inspreken van antwoorden sneller zou verlopen dan het intypen en om die reden is ervoor gekozen in de mondelinge variant de studenten alle vragen aan te bieden.

Net zoals beschreven in hoofdstuk 5 moesten de studenten zich voor de toets verplaatsen naar een computerlokaal dat dienst deed als toetsruimte.

Vooraf aan de toets werd een protocol doorlopen waarin de studenten mondeling werden ingeleid in de toets. Dit protocol is hetzelfde als beschreven in hoofdstuk 5 met enkele toevoegingen:

1. Na het inloggen in de toets kregen de studenten van elke toetstaak een oefenitem te maken. Deze toevoeging is hiervoor al gemotiveerd.
2. Na het afronden van hun toets werd het computerprogramma vervolgd met de hierboven beschreven protocolanalyse.

Data verzameling

De toetsen bestonden uit dezelfde items in dezelfde volgorden, waren identiek op de visuele representatievormen na maar verschilden in het instructiethema

waarmee de toets startte. Ter voorkoming dat bij elkaar zittende studenten bij elkaar zouden kunnen inkijken in het beantwoorden van de toetsvragen is er voor gekozen de toetsen te laten starten bij verschillende startitems van de instructiethema's. Met vijf instructiethema's zijn er ook vijf mogelijke startpunten waardoor de studenten, in het geval dat ze iets zouden zien van het beeldscherm bij een van de andere studenten, zo de indruk kregen een eigen toets te hebben. Elke toets bestond uit 57 CBT-items waarvan 19 items met herkenningstaken, 20 items met benoemingstaken en 18 items met relatietaken. Daarbij wordt opgemerkt dat de toets weliswaar bestaat uit 57 items maar in de nummering ook een item 58 aanwezig is en item 28 ontbreekt omdat deze laatste gedurende het ontwikkelproces is verwijderd en niet vervangen.

6.3 RESULTATEN

Overzicht resultaten algemene toets en toetsversies

Tabel 6.1a Overzicht resultaten van testanalyse van de algemene test en de toetsversies

	Algemene Test (N= 134)			FMC (N= 34)			FZC (N= 32)			SMC (N= 33)			SZC (N= 35)		
	<i>Gem</i>	<i>Var</i>	<i>Alpha</i>	<i>Gem</i>	<i>Var</i>	<i>Alpha</i>	<i>Gem</i>	<i>Var</i>	<i>Alpha</i>	<i>Gem</i>	<i>Var</i>	<i>Alpha</i>	<i>Gem</i>	<i>Var</i>	<i>Alpha</i>
T	0,67	0,020	0,85	0,63 ⁴	0,025	0,88	0,63 ⁴	0,021	0,85	0,68	0,020	0,87	0,72 ^{1,2}	0,012	0,78
H	0,79	0,018	0,59	0,75 ^{3,4}	0,026	0,70	0,75 ^{3,4}	0,025	0,70	0,82 ^{1,2}	0,010	0,28	0,83 ^{1,2}	0,007	0,09
B	0,68	0,039	0,80	0,65	0,050	0,82	0,66	0,036	0,74	0,69	0,038	0,83	0,73	0,031	0,77
R	0,52	0,032	0,69	0,47 ⁴	0,034	0,71	0,48 ⁴	0,027	0,63	0,53	0,042	0,78	0,59 ^{1,2}	0,020	0,52

Noot: T= Test (57 items); B= Benoemen (20 items); 1= Significant verschil met FMC, T-test; 3= Significant verschil met SMC, T-test; H= Herkennen (19 items); R= Relaties (18 items); 2= Significant verschil met FZC, T-test; 4= Significant verschil met SZC, T-test.

Tabel 6.1b *Vergelijking van de toetstaken per toetsversie*

Gepaarde T-toets			Gepaarde verschillen			t	df	Significantie (2-tailed)
Toets versie			<i>Gemiddeld</i>	<i>Standaard deviatie</i>	<i>Standaard meetfout</i>			
FMC	Paar 1	H - B	0,10	0,20	0,04	2,75	33	0,01
	Paar 2	H - R	0,28	0,16	0,03	10,56	33	≤0,01
	Paar 3	B - R	0,19	0,20	0,03	5,40	33	≤0,01
FZC	Paar 1	H - B	0,08	0,17	0,03	2,81	31	0,01
	Paar 2	H - R	0,26	0,16	0,03	9,47	31	≤0,01
	Paar 3	B - R	0,18	0,16	0,03	6,25	31	≤0,01
SMC	Paar 1	H - B	0,13	0,15	0,03	5,13	32	≤0,01
	Paar 2	H - R	0,29	0,19	0,03	8,96	32	≤0,01
	Paar 3	B - R	0,16	0,19	0,03	4,80	32	≤0,01
SZC	Paar 1	H - B	0,10	0,17	0,03	3,38	34	≤0,01
	Paar 2	H - R	0,24	0,12	0,02	11,40	34	≤0,01
	Paar 3	B - R	0,14	0,16	0,03	5,14	34	≤0,01

Tabel 6.1c Overzicht resultaten itemanalyse van algemene test en de toetsversies voor de toetstaak Herkennen

HERKENNEN		Algemene Test (N= 134)			FMC (N= 34)			FZC (N=32)			SMC (N= 33)			SZC (N= 35)		
Itemnr	Type	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr
1	H	0,87	0,12	0,265**	0,85	0,13	0,318	0,88	0,11	0,325	0,88	0,11	0,126	0,86	0,13	0,335*
3	H	0,79	0,17	0,352**	0,74	0,20	0,543**	0,84	0,14	0,138	0,76	0,19	0,336	0,83	0,15	0,346*
9	H	0,80	0,16	0,157	0,76	0,19	0,087	0,84	0,14	0,181	0,76	0,19	0,237	0,83	0,15	0,146
10	H	0,61	0,24	0,148	0,53	0,26	0,062	0,59	0,25	0,193	0,7	0,22	0,353*	0,63	0,24	-0,097
14	H	0,92	0,08	0,307**	0,88	0,11	0,342*	0,94	0,06	0,214	0,94	0,06	0,191	0,91	0,08	0,511**
15	H	0,90	0,09	0,329**	0,91	0,08	0,203	0,84	0,14	0,448*	0,88	0,11	0,315	0,94	0,06	0,336*
16	H	0,83	0,14	0,385**	0,74 ⁴	0,20	0,513**	0,81 ⁴	0,16	0,359*	0,79 ⁴	0,17	0,315	0,97 ^{1,2,3}	0,03	-0,063
17	H	0,91	0,08	0,161	0,82 ³	0,15	0,171	0,91	0,09	0,087	1 ¹	0		0,91	0,08	0,224
18	H	0,85 [#]	0,13	0,452**	0,82	0,15	0,215	0,69 ^{3, 4}	0,22	0,602**	0,94 ²	0,06	0,515**	0,94 ²	0,06	0,438**
19	H	0,80	0,16	0,421**	0,76	0,19	0,567**	0,69 ³	0,22	0,442*	0,91 ²	0,09	0,292	0,83	0,15	0,246
20	H	0,93	0,06	0,299**	0,94	0,06	0,064	0,84 ³	0,14	0,555**	1 ²	0		0,94	0,06	0,336*
25	H	0,40	0,24	0,317**	0,35	0,24	0,438**	0,38	0,24	0,423*	0,52	0,26	0,205	0,34	0,23	0,216
31	H	0,50	0,25	0,232**	0,53	0,26	0,242	0,5	0,26	0,481**	0,48	0,26	0,22	0,49	0,26	0,018
36	H	0,86	0,12	0,133	0,91	0,08	0,344*	0,88	0,11	0,419*	0,82	0,15	-0,055	0,83	0,15	-0,005
37	H	0,86	0,12	0,384**	0,79	0,17	0,630**	0,88	0,11	0,22	0,88	0,11	0,563**	0,89	0,10	-0,088
41	H	0,64 [#]	0,23	0,323**	0,53 ³	0,26	0,402*	0,41 ^{3, 4}	0,25	0,352*	0,82 ^{1, 2}	0,15	0,185	0,8 ²	0,17	0,005
49	H	0,70 [#]	0,21	0,281**	0,59 ⁴	0,25	0,161	0,59 ⁴	0,25	0,296	0,73	0,21	0,097	0,89 ^{1, 2}	0,10	0,476**
57	H	0,86	0,12	0,125	0,85	0,13	0,215	0,78	0,18	0,134	0,88	0,11	-0,04	0,91	0,08	0,039
58	H	0,93	0,07	0,034	0,94	0,06	0,503**	0,91	0,09	-0,139	0,91	0,09	-0,044	0,94	0,06	-0,273

Noot: (#)= Significant verschil $\leq 0,05$; ANOVA-test; (**) = Significantieniveau $\leq 0,01$; 1= Significant verschil met FMC, T-test; (*) = Significantieniveau $\leq 0,05$; 2= Significant verschil met FZC, T-test; **Zwakke items** = items met variantie $\leq 0,1$ of PB correlatie $\leq 0,2$; 3= Significant verschil met SMC, T-test; 4= Significant verschil met SZC, T-test.

Tabel 6.1d *Overzicht resultaten itemanalyse van algemene test en de toetsversies voor de toetstaak Benoemen*

BENOEMEN		Algemene Test (N= 134)			FMC (N= 34)			FZC (N= 32)			SMC (N= 33)			SZC (N= 35)		
Itemnr	Type	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr
2	B	0,39 #	0,24	0,137	0,44 ³	0,25	0,205	0,41 ³	0,25	0,068	0,09 ^{1,2,4}	0,09	0,071	0,6 ³	0,25	0,2
8	B	0,78	0,17	0,254**	0,76	0,19	0,087	0,75	0,19	0,426*	0,79	0,17	0,221	0,83	0,15	0,283
13	B	0,69	0,21	0,339**	0,65	0,24	0,301	0,69	0,22	0,317	0,64	0,24	0,449**	0,8	0,17	0,217
21	B	0,79	0,17	0,441**	0,68 ³	0,23	0,295	0,75	0,19	0,452**	0,88 ¹	0,11	0,540**	0,86	0,13	0,470**
23	B	0,74	0,19	0,420**	0,59 ³	0,25	0,459**	0,72	0,21	0,239	0,85 ¹	0,13	0,598**	0,8	0,17	0,3
24	B	0,78	0,17	0,478**	0,76	0,19	0,473**	0,75	0,19	0,587**	0,88	0,11	0,504**	0,74	0,20	0,442**
26	B	0,64 #	0,23	0,513**	0,5 ³	0,26	0,547**	0,56 ³	0,25	0,325	0,82 ^{1, 2}	0,15	0,505**	0,69	0,22	0,665**
27	B	0,69 #	0,22	0,465**	0,47 ^{3, 4}	0,26	0,666**	0,5 ^{3, 4}	0,26	0,217	0,85 ^{1, 2}	0,13	0,479**	0,91 ^{1, 2}	0,08	0,09
29	B	0,66	0,23	0,366**	0,65	0,24	0,294	0,66	0,23	0,433*	0,7	0,22	0,505**	0,63	0,24	0,313
30	B	0,55	0,25	0,430**	0,59	0,25	0,662**	0,47	0,26	0,421*	0,7 ⁴	0,22	0,253	0,46 ³	0,26	0,480**
32	B	0,70	0,21	0,427**	0,71	0,21	0,327	0,66	0,23	0,621**	0,58 ⁴	0,25	0,552**	0,86 ³	0,13	0,065
33	B	0,78	0,17	0,216*	0,82	0,15	0,198	0,66	0,23	0,131	0,85	0,13	0,447**	0,80	0,17	0,087
38	B	0,70 #	0,21	0,307**	0,76 ³	0,19	0,268	0,72	0,21	0,429*	0,52 ^{1, 4}	0,26	0,282	0,80 ³	0,17	0,382*
39	B	0,85	0,13	0,407**	0,82	0,15	0,259	0,78	0,18	0,453**	0,88	0,11	0,480**	0,91	0,08	0,410*
40	B	0,58	0,25	0,485**	0,65	0,24	0,469**	0,41	0,25	0,423*	0,64	0,24	0,481**	0,63	0,24	0,635**
42	B	0,71 #	0,21	0,381**	0,53 ³	0,26	0,322	0,69	0,22	0,233	0,88 ¹	0,11	0,445**	0,74	0,20	0,528**
43	B	0,75	0,19	0,516**	0,62 ³	0,24	0,576**	0,75	0,19	0,488**	0,85 ¹	0,13	0,436*	0,77	0,18	0,530**
44	B	0,80	0,16	0,512**	0,71	0,21	0,546**	0,88	0,11	0,478**	0,73	0,21	0,651**	0,89	0,10	0,313
45	B	0,75	0,19	0,546**	0,68	0,23	0,580**	0,81	0,16	0,28	0,7	0,22	0,664**	0,8	0,17	0,724**
56	B	0,35 #	0,23	0,022	0,71 ^{3, 4}	0,21	0,451**	0,69 ^{3, 4}	0,22	0,317	0 ^{1, 2}	0		0,03 ^{1, 2}	0,03	0,006

Noot: (#)= Significant verschil $\leq 0,05$; ANOVA-test; (**)= Significantieniveau $\leq 0,01$; 1= Significant verschil met FMC, T-test; (*)= Significantieniveau $\leq 0,05$; 2= Significant verschil met FZC, T-test; **Zwakke items** = items met variantie $\leq 0,1$ of PB correlatie $\leq 0,2$; 3= Significant verschil met SMC, T-test; 4= Significant verschil met SZC, T-test.

Tabel 6.1e Overzicht resultaten itemanalyse van algemene test en de toetsversies voor de toetstaak Relaties

RELATIES		Algemene Test (N= 134)			FMC (N= 34)			FZC (N= 32)			SMC (N= 33)			SZC (N= 35)		
Itemnr	Type	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr	Gem	Var	PB corr
4	R	0,87	0,11	0,381**	0,82 ⁴	0,15	0,346*	0,91	0,09	0,367*	0,79 ⁴	0,17	0,570**	0,97 ^{1, 3}	0,03	-0,006
5	R	0,84	0,13	0,230**	0,82	0,15	0,469**	0,91	0,09	0,18	0,79	0,17	0,154	0,86	0,13	0,119
6	R	0,44	0,25	0,197*	0,53	0,26	0,235	0,5	0,26	0,202	0,33	0,23	0,393*	0,40	0,25	0,127
7	R	0,31	0,21	0,136	0,35	0,24	0,103	0,34	0,23	0,179	0,33	0,23	0,278	0,20	0,17	0,137
11	R	0,21	0,17	0,239**	0,18	0,15	0,345*	0,28	0,21	0,331	0,18	0,15	0,156	0,20	0,17	0,172
12	R	0,34	0,23	0,405**	0,29	0,21	0,537**	0,31	0,22	0,495**	0,33	0,23	0,393*	0,43	0,25	0,118
22	R	0,40	0,24	0,366**	0,32	0,23	0,453**	0,41	0,25	0,273	0,45	0,26	0,607**	0,43	0,25	0,061
34	R	0,81	0,16	0,300**	0,79	0,17	0,481**	0,81	0,16	0,26	0,85	0,13	0,383*	0,77	0,18	0,103
35	R	0,49	0,25	0,362**	0,56	0,25	0,339*	0,44	0,25	0,480**	0,45	0,26	0,522**	0,49	0,26	0,178
46	R	0,57 #	0,25	0,442**	0,32 ^{3, 4}	0,23	0,218	0,38 ^{3, 4}	0,24	0,487**	0,76 ^{1, 2}	0,19	0,390*	0,83 ^{1, 2}	0,15	0,496**
47	R	0,46 #	0,25	0,424**	0,32 ³	0,23	0,361*	0,34 ⁴	0,23	0,367*	0,52	0,26	0,406*	0,66 ^{1, 2}	0,23	0,410*
48	R	0,57 #	0,25	0,472**	0,32 ^{3, 4}	0,23	0,425*	0,38 ^{3, 4}	0,24	0,335	0,73 ^{1, 2}	0,21	0,461**	0,83 ^{1, 2}	0,15	0,458**
50	R	0,50	0,25	0,373**	0,41	0,25	0,185	0,5	0,26	0,551**	0,45	0,26	0,251	0,63	0,24	0,508**
51	R	0,61	0,24	0,230**	0,56	0,25	0,225	0,66	0,23	0,099	0,58	0,25	0,263	0,66	0,23	0,360*
52	R	0,81	0,16	0,308**	0,79	0,17	0,291	0,84	0,14	0,149	0,82	0,15	0,435*	0,77	0,18	0,474**
53	R	0,58	0,25	0,536**	0,47	0,26	0,466**	0,53	0,26	0,473**	0,64	0,24	0,513**	0,69	0,22	0,676**
54	R	0,48 #	0,25	0,398**	0,53 ²	0,26	0,503**	0,13 ^{1, 3, 4}	0,11	0,108	0,55 ²	0,26	0,400*	0,69 ²	0,22	0,391*
55	R	0,03	0,03	-0,026	0	0		0,06	0,06	-0,086	0	0		0,06	0,06	-0,031

Noot: (#)= Significant verschil $\leq 0,05$; ANOVA-test; (**) = Significantieniveau $\leq 0,01$; 1= Significant verschil met FMC, T-test; (*) = Significantieniveau $\leq 0,05$; 2= Significant verschil met FZC, T-test; **Zwakke items** = items met variantie $\leq 0,1$ of PB correlatie $\leq 0,2$; 3= Significant verschil met SMC, T-test; 4= Significant verschil met SZC, T-test.

Moeilijkheidsgraad en betrouwbaarheid van de toetsen

In tabel 6.1a is van de algemene toets en de verschillende toetsversies de gemiddelde testscore, testvariantie en de betrouwbaarheid zoals geschat met behulp van Cronbach's alpha weergegeven. Uit de tabel blijkt dat de toets als geheel een betrouwbaarheid heeft van 0,85, hetgeen ruim groter is dan de ondergrenswaarde ($\geq 0,7$). Dit geldt ook voor de gehele toets per toetsversie (FMC, FZC, SMC en SZC) met respectievelijk de betrouwbaarheid $\alpha_{FMC} = 0,88$; $\alpha_{FZC} = 0,85$; $\alpha_{SMC} = 0,87$; $\alpha_{SZC} = 0,78$. De algemene toets en de vier varianten kunnen met deze waarden als betrouwbaar worden beschouwd, daar de betrouwbaarheid een afspiegeling is van het percentage van de variantie (α -waarde * 100%) dat minimaal toe te schrijven is aan de ware score.

Uit tabel 6.1a blijkt dat per toetstaak de betrouwbaarheid niet voor alle toetsversies voldoet aan de ondergrenswaarde ($\geq 0,7$). Voor de toetstaak herkennen voldoet de betrouwbaarheid niet voor de toets als geheel ($\alpha_H = 0,59$) en niet voor de toetsversies met schema's: $\alpha_{SMC\_H} = 0,28$ en $\alpha_{SZC\_H} = 0,09$. De lage betrouwbaarheid van de algemene toets blijkt vooral te worden veroorzaakt door de lage betrouwbaarheid van de schematische toetsversies waarvan minimaal 28% respectievelijk 9% van de variantie is toe te wijzen is aan de variantie van de ware score. Voor het toetsen van dit deelconstruct zijn alleen de toetsversies met foto's geschikt en dienen de items voor de schematische vormen te worden verbeterd.

Voor de toetstaak benoemen voldoen de algemene toets en de vier varianten ruim aan de betrouwbaarheidseisen en zijn alle vier varianten geschikt voor het toetsen van het kunnen benoemen van anatomische objecten.

Voor de toetstaak relaties is voor de algehele toets de betrouwbaarheid 0,69. Hoewel strikt genomen kleiner dan 0,7, wordt dit betrouwbaarheidsniveau hier toch als voldoende beschouwd. Voor deze toetstaak voldoen voor de toetsversies zonder context de betrouwbaarheidswaarden niet aan de ondergrenswaarde, $\alpha_{FZC\_R} = 0,63$ en $\alpha_{SZC\_R} = 0,52$, hetgeen de lage betrouwbaarheid van de algemene toets op dit onderdeel verklaart. De toetsversies met context blijken geschikt voor toetsing van de functionele relaties tussen anatomische objecten, terwijl de versies zonder context verbeterd zouden moeten worden.

Bij vergelijking van de toetsversies op de T-toetsen, tabel 6.1 a, zijn er significante verschillen tussen de algemene toetsscores van FMC, FZC en SZC. De toetsversie SZC heeft een significant hoger gemiddelde testscore dan de toetsen in fotovorm.

Voor de toetstaak herkennen hebben de toetsversies met schema's significant hogere gemiddelden dan de toetsversies met foto's. Voor de toetstaak met relaties is er een significant verschil tussen de toetsversies met foto's en de toetsversie SZC in het voordeel van SZC.

Bij vergelijking van de toetstaken in elke toetsversie, tabel 6.1b, blijken de gemiddelde toetsscores van elkaar te verschillen. Deze verschillen zijn in alle toetsversies significant waarmee mag worden geconcludeerd dat er sprake is van een moeilijkheidsgraad. Deze is van makkelijk naar moeilijk ofwel herkennen is het makkelijkst, benoemen moeilijker en relaties het moeilijkst.

Deze eerste resultaten over de toetsversies en de toetstaken leiden gecombineerd tot de volgende bevinding: de toetsversies zijn elk voldoende betrouwbaar. Bij uitsplitsing naar de toetstaken en daar waar de verschillen tussen de versies significant is, blijkt dat toetstaken met een hoger gemiddelde een lagere betrouwbaarheid hebben die niet voldoet aan de betrouwbaarheidsondergrens in tegenstelling tot de toetsversies waar een significant verschil mee is die weliswaar een lager gemiddelde hebben maar waarvan de betrouwbaarheid wel voldoende is.

De resultaten tot nu suggereren dat voor het herkennen van anatomische objecten de toetsversies met foto's beter functioneren en voor het leggen van functionele relaties de toetsversies met context.

Om meer inzicht te krijgen in de kwaliteit van de items zijn deze verder geanalyseerd op itemniveau voor elke toetstaak en representatievorm. In tabel 6.1c staan de resultaten van de itemanalyse weergegeven voor de *herkenningstaak* van de algemene toets en de vier representatievormen.

Uit de resultaten voor de herkenningstaak blijkt dat:

1. er zwakke items zijn in:
  - a. de algemene test voor de items 9,10, 14, 15, 17, 20, 36, 57 en 58;
  - b. foto met context voor de items items 9, 10, 15, 17, 20, 36, 49 en 58;
  - c. foto zonder context voor de items 3, 9, 10,14,17, 57 en 58
  - d. schema met context voor de items 1, 14, 17, 18, 19, 20, 36, 41, 49, 57 en 58
  - e. schema zonder context voor de items 9, 10, 14, 15, 16, 17, 18, 20, 31, 36, 37, 41, 49, 57 en 58
2. op basis van een ANOVA-analyse er significante verschillen zijn voor hetzelfde item in verschillende toetsversies voor de items 18, 41 en 49. Op

basis van de ANOVA zijn er geen significante verschillen gevonden voor de items 16, 17, 19 en 20 terwijl bij de onderlinge vergelijking van de toetsversies met T-testen er wel significante verschillen blijken te zijn.

3. op basis van T-testen naar het itemgedrag in de verschillende toetsversies blijkt dat
  - a. er significante verschillen zijn voor item 16, het zwakke item in SZC-vorm, waarbij FMC, FZC en SMC significant verschillen met SZC ten voordele van SZC
  - b. er significante verschillen zijn voor de items 17 en 20, het zwakke item in SMC-vorm ten opzichte van een toetsversie met foto (FMC danwel FZC). In de SMC vorm heeft een ieder deze items goed.
  - c. er significante verschillen zijn voor item 18 in het voordeel van beide schematische toetsversies ten opzichte van FZC
  - d. er significante verschillen zijn voor items 41 en 49 waarbij de schematische vormen significant hoger scoren dan de toetsversies met foto's
4. de zwakke items voornamelijk items zijn met een hoog gemiddelde, die gemakkelijk zijn en daarmee niet onderscheidend. Deze items zijn onvoldoende informatief voor het construct dat gemeten wordt en bevinden zich vooral in de schematische toetsversies. Dat verklaart de lage betrouwbaarheidswaarden voor deze toetsversies.

In tabel 6.1d staan de resultaten van de itemanalyse weergegeven voor de *benoemingstaak* van de algemene toets en de vier representatievormen.

Uit de resultaten voor de benoemingstaak blijkt dat:

1. er zwakke items zijn in:
  - a. de algemene test voor de items 2 en 56
  - b. de toetsversie foto met context voor de items 8 en 33
  - c. de toetsversie foto zonder context voor de items 2 en 33
  - d. de toetsversie schema met context voor de items 2 en 56
  - e. de toetsversie schema zonder context voor de items 2, 27, 32, 33, 39, 44 en 56
2. op basis van een ANOVA-analyse er significante verschillen zijn voor hetzelfde item in verschillende toetsversies voor de items 2, 26, 27, 38, 42 en 56. Op basis van de ANOVA zijn er geen significante verschillen gevonden voor de items 21, 23, 30, 32 en 43 terwijl bij de onderlinge vergelijking van de toetsversies met T-testen er wel significante verschillen blijken te zijn.
3. op basis van T-testen naar het itemgedrag in de verschillende toetsversies blijkt dat:

- a. er significante verschillen zijn voor item 2 waarbij de toetsversies met foto's gemiddeld hoger scoren dan de schematische vormen met als laagste het gemiddelde voor SMC
  - b. er significante verschillen zijn voor items 21, 23, 42 en 43 voor de toetsversies FMC en SMC, in het voordeel van SMC
  - c. item 26 in de fotovormen significant lager is dan in SMC
  - d. item 27 in de foto toetsversies significant lager scoort dan in de schematische vormen
  - e. item 30 in de toetsversie SMC significant hoger scoort dan in SZC
  - f. item 32 in de toetsversie SZC significant hoger scoort dan in SMC
  - g. item 38 significant hoger scoort voor FMC, SZC dan in SMC
  - h. item 56 significant verschilt tussen de fotovormen die hoger liggen en de schematische vormen. Dit item lijkt te moeilijk te zijn gegeven de erg lage gemiddelde itemscore en variantie.
4. Er weinig zwakke items zijn in de algemene toets en de toetsversies met uitzondering van de toetsversie schema zonder context. Deze toetsvorm blijkt ook voor deze toetstaak zwak met een meerderheid aan zwakke items.

In tabel 6.1e staan de resultaten van de itemanalyse weergegeven voor de *relatietaak* van de algemene toets en de vier representatievormen. Uit de resultaten voor de *realtietaak* blijkt dat:

1. er zwakke items zijn in:
  - a. de algemene test voor de items 6, 7 en 55
  - b. de toetsvorm foto met context voor de items 7, 50 en 55
  - c. de toetsvorm foto zonder context voor de items 4, 5, 7 51, 52, 54 en 55
  - d. de toetsvorm schema met context voor de items 5, 11, en 55
  - e. de toetsvorm schema zonder context voor de items 4, 5, 6, 7, 11, 12, 22, 34, 35 en 55
2. op basis van ANOVA-analyse er significante verschillen zijn voor hetzelfde item in verschillende toetsversies voor de items 46, 47, 48 en 54. Op basis van de ANOVA zijn er geen significante verschillen gevonden voor item 4 terwijl bij de onderlinge vergelijking van de toetsversies met T-testen er wel significante verschillen blijken te zijn.
3. op basis van T-testen naar het itemgedrag in de verschillende toetsversies blijkt dat:
  - a. er significante verschillen zijn voor item 4 waarbij de toetsversie SZC gemiddeld hoger scoort dan de toetsversies met context.

- b. er zijn significante verschillen zijn voor items 46 en 48 waarbij de schematische vormen hoger scoren dan toetsversies met foto's
  - c. item 47 in de fotovormen significant lager scoort dan SZC
  - d. item 54 significant hoger scoort voor FMC, SMC, SZC dan voor FZC.
4. er een beperkt aantal zwakke items is in de algemene toets en in de toetsversies, met uitzondering van de toetsversie schema zonder context die de meeste zwakke items heeft. De toetsversies met foto's en de versie schema met context lijken afdoende items te hebben die onderscheidend zijn voor het meten van het geleerde deelconstruct: het kunnen leggen van functionele relaties tussen anatomische onderdelen.

Op basis van de analyses van de eerste ontwikkelde CBT-toets in hoofdstuk 5 en van de tweede doorontwikkelde CBT-toets in het hier gepresenteerde experiment, zijn beide versies met elkaar vergeleken. In tabel 6.2 zijn de aantallen items die als zwak zijn getypeerd weergegeven voor beide versies per taak voor zowel de algemene toets als de vier representatievormen van de toets. Zo ook zijn de aantallen items die voldoende zijn bevonden weergegeven met de bijbehorende percentages.

Tabel 6.2 *Vergelijking 1^{ste} CBT toetsversie met de 2^{de} verbeterde CBT toetsversie*

Herkennen		Test		FMC		FZC		SMC		SZC	
Items ↓		Versie 1	Versie 2	Versie 1	Versie 2	Versie 1	Versie 2	Versie 1	Versie 2	Versie 1	Versie 2
Aantal	N	20	19	20	19	20	19	20	19	20	19
Niet zwak	N	5	10	8	11	7	12	11	8	5	4
	%	25	53	40	58	35	63	55	42	25	21
Zwak	N	15	9	12	8	13	7	9	11	15	15

Benoemen		Test		FMC		FZC		SMC		SZC	
Items ↓		Versie 1	Versie 2	Versie 1	Versie 2	Versie 1	Versie 2	Versie 1	Versie 2	Versie 1	Versie 2
Aantal	N	21	20	21	20	21	20	21	20	21	20
Niet zwak	N	19	18	19	18	18	18	18	18	13	13
	%	90	90	90	90	86	90	86	90	62	65
Zwak	N	2	2	2	2	3	2	3	2	8	7

Relaties		Test		FMC		FZC		SMC		SZC	
Items ↓		Versie 1	Versie 2	Versie 1	Versie 2	Versie 1	Versie 2	Versie 1	Versie 2	Versie 1	Versie 2
Aantal	N	14	18	14	18	14	18	14	18	14	18
Niet zwak	N	10	15	7	15	7	11	7	15	7	8
	%	71	83	50	83	50	61	50	83	50	44
Zwak	N	4	3	7	3	7	7	7	3	7	10

Voor wat betreft de kwaliteit van de toets is de conclusie dat de toets in drie representatievormen -FMC, FZC en SMC- is verbeterd ten opzichte van de vorige versie waarbij het onderdeel herkennen in beide schematische toetsversies zwak blijft.

De toetsversie SZC is niet verbeterd in de huidige versie. Voor de toetstaak herkennen zijn de toetsversies met foto's verbeterd en bevatten het minste aantal zwakke items in vergelijking tot de toetsversies met schema's.

Naar aanleiding van de gepresenteerde resultaten zijn aansluitende conclusies:

- Voor de toetsen:
  - Dat zowel de algemene toets als de vier toetsversies een betrouwbaarheidswaarde ruim boven de ondergrens hebben
  - Dat er een significant verschil is tussen de gemiddelde toetsscores van de toetsversies met foto's en de toetsversie schema zonder context in het voordeel van de laatst genoemde
  - Dat er sprake is van een moeilijkheidsgraad met een patroon dat zich voortdoet in elke toetsversie
- Voor de toetstaak herkennen:
  - dat toetsversies met foto's voldoen aan de norm voor de ondergrens van betrouwbaarheid in tegenstelling tot de toetsversies met schema's die ruim onder de maat zijn
  - dat toetsversies met schema's hebben significant hogere gemiddelden dan de toetsversies met foto's
- Voor de toetstaak benoemen:
  - dat alle toetsversies voldoen voor dit onderdeel ruim aan de betrouwbaarheidsnorm
  - dat deze toetstaak ook na aanpassing van de taak nog steeds een stabiel toetsonderdeel is met een kleine verbetering
- Voor de toetstaak relaties:
  - dat de toetsversies foto zonder context en schema zonder context niet voldoen aan de norm voor de betrouwbaarheid, in tegenstelling tot de toetsversies met context die wel aan de norm voldoen
  - dat er een significant verschil is tussen de toetsversies met foto's en de toetsversie SZC in het voordeel van SZC
  - dat er voor vijf items een significant verschil is tussen twee of meer toetsversies

Invloed van realisme en context

De invloed van de variabelen realisme en context is onderzocht met een ANOVA-analyse uitgevoerd voor de gehele test. Uit deze analyse blijkt er een hoofdeffect te zijn voor de variabele realisme en niet voor de variabele context. Er is geen sprake van een interactie-effect tussen realisme en context. In tabel 6.3 zijn de gemiddelde scores weergegeven voor de variabelen op foto, schema, met context en zonder context. Daarin is te zien dat het gevonden significante effect in het voordeel is van schema.

Tabel 6.3 ANOVA: overzicht effecten variabelen op test

		Realisme #		
		Foto	Schema	Totaal
Context	Met context	0,63	0,68	0,66
	Zonder context	0,64	0,72	0,682
	Totaal	0,63 ¹	0,70 ²	

Noot: # = ANOVA, hoofdeffect; significantie niveau $\leq 0,01$; 1 = significant verschil met schema; 2 = significant verschil met foto.

Vervolgens is een MANOVA -analyse op de verschillende toetstaken afzonderlijk uitgevoerd. Daaruit blijkt dat er significante hoofdeffecten zijn voor realisme in de toetstaken herkennen en relaties. Er zijn verder geen significante hoofdeffecten gevonden voor realisme, ook niet voor context op de toetstaken. Ook zijn er geen interactie-effecten gevonden tussen de variabelen op de toetstaken. Tabel 6.4 geeft een overzicht van de effecten van de variabelen op het taakniveau.

Tabellen 6.3 en 6.4 geven weer dat de gevonden significante effecten in beide gevallen in het voordeel zijn van schema.

Voor de toetstaak herkennen moet het gevonden significant hoofdeffect in het voordeel van schema's worden gezien in relatie tot de eerdere analyseresultaten op deze taak. De toetsversies met schema's hebben weliswaar significant hogere gemiddelden dan de toetsversies met foto's maar tegelijkertijd hebben deze een betrouwbaarheidswaarde die ruim onder de norm ligt. De toetsversies met schema's bevatten ook de zwakke items.

Tabel 6.4 MANOVA: overzicht effecten op de toetstaken

			Gemiddelde	Standaard deviatie	N
Herkennen	Realisme #	Foto	0,75 ¹	0,16	66
		Schema	0,82 ²	0,09	68
	Context	Met context	0,79	0,14	67
		Zonder context	0,79	0,13	67
Benoemen	Realisme	Foto	0,66	0,21	66
		Schema	0,71	0,19	68
	Context	Met context	0,67	0,21	67
		Zonder context	0,70	0,18	67
Relaties	Realisme #	Foto	0,48 ¹	0,17	66
		Schema	0,56 ²	0,18	68
	Context	Met context	0,50	0,20	67
		Zonder context	0,54	0,16	67

Noot: # = MANOVA, hoofdeffect; significantie niveau $\leq 0,01$; 1 = significant verschil met schema; 2 = significant verschil met foto.

Voor de toetstaak benoemen zijn er geen significante hoofdeffecten gevonden. Dit resultaat is aannemelijk omdat in de eerdere analyses ook geen significante verschillen tussen de toetsversies zijn gevonden en alle toetsversies voldoen aan de betrouwbaarheidsnorm. Op itemniveau zijn er significante verschillen gevonden tussen de toetsversies maar deze variëren soms in het voordeel van foto's en soms in het voordeel van schema's.

Voor de toetstaak relaties is het gevonden hoofdeffect in het voordeel van schema. Ook hier moet het verkregen resultaat worden gezien in relatie tot eerdere analyseresultaten. De algemene test voor deze toetstaak voldoet eigenlijk net niet aan de betrouwbaarheidsnorm, maar met een betrouwbaarheid van 0,69 kan de algemene test toch als voldoende worden beschouwd. Dat is anders voor de toetsversies zonder context waarvan de alpha's duidelijk onder de betrouwbaarheidsgrens liggen. Enkel de toetsversies met context voldoen aan de norm en daarvan zijn er drie items die significant verschillen in beide toetsversies waarvan twee in het voordeel van schema.

Relatie tussen de constructen in de verschillende toetsversies

In de vorige paragrafen is de betrouwbaarheid en de moeilijkheidsgraad van de hele toets en de drie toetstaken voor de deelconstructen Herkennen, Benoemen en Relaties onderzocht. Daarmee is nog niet de vraag beantwoord of de vier toetsversies (FMC, FZC, SMC en SZC) ook vergelijkbare constructen meten. Om

daar enig inzicht in te krijgen is geanalyseerd of de relatie tussen de drie constructen en de eigenschappen van de items in alle vier de toetsversies vergelijkbaar zijn. Als bijproduct van deze analyses wordt extra informatie geleverd over de betrouwbaarheid van de toetsversies en de bijdragen van de individuele items daaraan. De analyses zijn uitgevoerd met item response theorie (IRT, van der Linden & Hambleton, 1997). Deze techniek is nauw verwant aan factoranalyse en staat daarom ook wel bekend als ‘full-information factor analysis’ (Ackerman, 1994, 1996).

In de kern is IRT de kans op een goed antwoord van een student n , op een item i , afhankelijk van een vaardigheidsparameter θ_n en twee item parameters a_i en b_i . Daarbij is a_i de zogenaamde discriminatieparameter en b_i de moeilijkheidsparameter. De kans op een goed antwoord voor twee items is weergegeven in onderstaand figuur 6.3. Op de x-as staat de vaardigheid θ_n . Op de y-as staat de kans op een goed antwoord die stijgt als het vaardigheidsniveau stijgt. De moeilijkheidsparameter b_i bepaalt de positie van het item op de vaardigheidsschaal. De discriminatieparameter a_i bepaalt hoe stijl de curve van het item is. Hoe hoger a_i , hoe betrouwbaarder het item is. Een van de belangrijke voordelen van IRT is dat als niet iedereen precies dezelfde items heeft gemaakt, en dus de totaal scores niet meer vergelijkbaar zijn, de vaardigheidsscores θ_n dat wel zijn.

Figuur 6.3 Voorbeeld van de vaardigheid op twee items, i_1 en i_2

Alle analyses werden uitgevoerd met het public domain software pakket MIRT (2010). Het programma berekent marginale maximum likelihood schattingen voor multidimensionale IRT modellen, en chi-kwadraat model-passingstoetsen. De in het programma geïmplementeerde methoden zijn onder andere verantwoord in Glas (1999, 2010).

Omdat, zoals uit de hierna gerapporteerde analyses duidelijk zal worden, de items sterk verschillen in discriminatie parameters is er niet voor gekozen om het Rasch model (Rasch, 1960) te gebruiken. Voor het gebruik van het Rasch model is het uitgangspunt dat alle discriminatieparameters gelijk zijn.

Voor alle analyses is voor iedere toetstaak een apart IRT-model gedefinieerd. Elke student heeft drie vaardigheidsparameters, één voor Herkennen, één voor Benoemen en één voor Relaties. Dus de drie constructen, Herkennen, Benoemen en Relaties, worden vertaald in drie IRT-vaardigheidsparameters, c.q. in drie dimensies. Tenslotte hebben de drie vaardigheidsparameters een correlatiematrix. Concreet worden twee vragen onderzocht:

1. Zijn de correlatiematrices van de toetstaken Herkennen, Benoemen en Relaties in de vier toetsversies vergelijkbaar.
2. Zijn de itemparameters in de vier populaties met elkaar vergelijkbaar.

Als het antwoord op beide vragen ja is, is het aannemelijk dat de vier versies vergelijkbare constructen meten omdat de correlaties tussen de drie dimensies in de vier versies hetzelfde zijn, en omdat de items in de vier versies dezelfde responsekansen hebben. Voor een IRT-analyse worden alle data van alle items en alle toetsversies samengevoegd om een IRT-model te schatten. De werkwijze bij deze analyses is het zoeken van het best passende model hetgeen een iteratief proces is, omdat het model zowel itemparameters als één of meer correlatiematrices bevat.

De passing van de IRT-modellen bij de data werd geëvalueerd met Akaike's informatie criterium (AIC), tabel 6.5. De reden hiervoor is als volgt: analoog aan de hypothesetoetsing over correlatiematrices in hoofdstuk 5, kan men modellen met elkaar vergelijken met de "Log-likelihood", c.q. de "deviance". Te zien is, in de tweede kolom, dat de hoogste deviance overeenkomt met het laagste aantal vrijheidsgraden in het model(het aantal te schatten parameters) en dat de laagste deviance overeenkomt met het hoogste aantal vrijheidsgraden. Echter, vergeleken met de modellen in hoofdstuk 5, heeft een IRT-model erg veel parameters. Daarom bevoordeelt deze aanpak modellen met veel parameters en leidt dit tot "overfitten". De AIC is een deviance met een correctie voor het aantal parameters. Modellen worden onderling vergeleken met behulp van de AIC en het informatiefste model is het model met de kleinste AIC.

Tabel 6.5 geeft door middel van een selectie van analyses een beknopt overzicht van de reeks analyses die is uitgevoerd. Allereerst de analysemodellen één tot en met vier waarin is nagegaan of en welke correlatiematrices in een model bij elkaar passen. De itemparameters werden over de toetsversies per item gelijk verondersteld. In analyse 1 is er voor iedere toetsversie een unieke correlatiematrix tussen de drie dimensies gedefinieerd. De vier geschatte matrices waren vergelijkbaar (steeds hoge positieve correlaties tussen de dimensies), maar niet exact gelijk. Om een eerste indruk te krijgen van het verschil tussen de toetsversies werd het aantal matrices verkleind. Analoog aan de hypothesetoetsing over correlatiematrices in hoofdstuk 5, werden voor alle 15 combinaties van correlatiematrices de deviances berekend. Op basis van de gevonden AIC werden drie oplossingen als uitgangspunt voor de verdere analyses gekozen. In de eerste oplossing werden de matrices voor FMC en FZC gelijk verondersteld (analyse 2). In de tweede oplossing werden de matrices voor FMC, FZC en SZC gelijk verondersteld (analyse 3) en de derde oplossing veronderstelde dezelfde matrix voor alle toetsversies (analyse 4).

Uit de eerste vier modellen blijkt, in tabel 6.5, dat model 2 met een gelijke correlatie matrix voor FMC en FZC, één voor SMC, en één voor SZC het best bij de data past. Dit echter nog steeds onder de veronderstelling dat alle items voor alle toetsversies hetzelfde functioneren, c.q. dat ze in alle toetsversies vergelijkbare discriminatieparameter (a_i) en moeilijkheidsparameter (b_i) hebben.

Tabel 6.5 *IRT analyses*

Analyse Model	Deviance	AIC	Vrijheids-graden	Aantal correlatie matrices	Aantal items verschillend over toetsversies
1	7638,136	7926,136	144	4	0
2	7641,058	7911,058	135	3	0
3	7661,552	7913,552	126	2	0
4	7687,464	7921,464	117	1	0
5	7284,834	7622,834	169	3	17
6	7308,458	7628,458	160	2	17

Daarna werd de mate waarin items in het model passen geëvalueerd door voor ieder item een “item-fit statistic” uit te rekenen. Het gaat hierbij om een chi-kwadraat die gebaseerd is op het verschil tussen de verwachte en geobserveerde gemiddelde itemscores in iedere toetsversie. Een significante uitkomst van de chi-

kwadraattoets is een indicatie dat het item slecht in het model past. Verder is ook van belang hoe ver de geobserveerde en verwachte gemiddelden uit elkaar liggen. Dit wordt uitgedrukt door een zogenaamde “effect size”. Het criterium hierbij is dat de effect size kleiner moet zijn dan 0,10. Uit de item-fit-statistics uitgerekend op de modellen 2 en 3 met de kleinste AIC waarden, bleek dat deze niet houdbaar waren omdat voor een aantal items de fit-toetsen significant waren en de effect sizes te groot. Om die reden werden de analyses vervolgd.

In een aantal stappen werden steeds in de volgende analyses de items met de grootste effect sizes van unieke itemparameters voor specifieke toetsversies voorzien. Met andere woorden, voor deze items wordt de veronderstelling ingevoerd dat hun discriminatie a_i en moeilijkheidsparameter b_i over de toetsversies verschilt. Voor de overige items blijft dus de veronderstelling gelden dat hun discriminatie a_i en moeilijkheidsparameter b_i niet over de toetsversies verschilt. Na een iteratief analyseproces zijn analyses 5 en 6 het eindresultaat. Na het toevoegen van 34 itemparameters voor 17 items (per item een discriminatie- en een moeilijkheidsparameter) ontstond een zeer goed passend model: er waren geen significante (significantieniveau 5%) item-fit toetsen meer over en alle effect sizes waren onder de 0,10. Uit de AIC-waarden van de modellen 5 en 6 blijkt vervolgens het model met drie correlatiematrices het beste te passen omdat dit model de laagste AIC waarde heeft. Dit is model 5 met een gelijke correlatie matrix voor FMC en FZC, één voor SMC, en één voor SZC, gelijk model 2 maar waarvan 17 items verschillend functioneren in de toetsversies.

Zoals hierboven al aangegeven is de vraag of de vier toetsversies dezelfde constructen meten, te onderzoeken door de correlatiestructuur en de structuur van de itemparameters van de vier toetsversies te onderzoeken. Als de correlatiestructuur en de itemparameters in de vier toetsversies perfect gelijk zijn, meten de vier toetsversies exact hetzelfde construct. Als er verschillen in de correlaties en itemparameters tussen de toetsversies zijn, duidt dit op een bepaalde mate van verschillen in de gemeten constructen.

De drie geschatte correlatiematrices staan in tabel 6.6. Alle correlaties zijn positief en boven de 0,50. Hoewel de drie matrices op details verschillen en de hypothese dat ze gelijk zijn statistisch verworpen moet worden, zijn ze goed vergelijkbaar en geven geen indicatie dat de drie constructen in de verschillende toetsversies sterk verschillend functioneren.

Als we de resultaten in tabel 6.6 vergelijken met de resultaten uit de tabellen 5.3 en 5.4 zien we twee aspecten:

1. Hier is de correlatiestructuur van FMC en FZC vergelijkbaar, terwijl in hoofdstuk 5 de correlatiestructuur van FZC en SZC vergelijkbaar zijn. Dit is te verklaren door de wijzigingen in de toets.
2. De correlaties in Tabel 6.6 zijn hoger dan in Tabel 5.3. Dit is een bekend verschijnsel: de correlaties in Tabel 6.6 zijn correlaties tussen “ware scores” (gecorrigeerd voor onbetrouwbaarheid) en de correlaties in Tabel 5.3 zijn correlaties tussen “geobserveerde scores” die verlaagd zijn door onbetrouwbaarheid (attenuatie-effect).

Tabel 6.6 *Correlatiematrices van latente variabelen Herkennen, Benoemen en Relaties*

FMC en FZC			
	Herkennen	Benoemen	Relaties
Herkennen	1,000	0,749	0,709
Benoemen	0,749	1,000	0,832
Relaties	0,709	0,832	1,000
SMC			
	Herkennen	Benoemen	Relaties
Herkennen	1,000	0,992	0,574
Benoemen	0,992	1,000	0,674
Relaties	0,574	0,674	1,000
SZC			
	Herkennen	Benoemen	Relaties
Herkennen	1,000	0,765	0,956
Benoemen	0,765	1,000	0,695
Relaties	0,956	0,695	1,000

Verdere evidentie voor de meting van dezelfde constructen in de verschillende toetsversies is te vinden in de tabellen 6.7a, 6.7b en 6.7c waar de schattingen van de item parameters zijn weergegeven. In de kolommen met de labels FMC, FZC, SMC en SZC staat een 1 als de itemparameter betrekking heeft op de genoemde toetsversie. Items met specifieke parameters voor specifieke toetsversies zijn zo eenvoudig te herkennen. Bijvoorbeeld, in Tabel 6.7a staan twee rijen voor item 16, de eerste bevat de item parameters voor de toetsversies FMC, FZC, en SMC, en de tweede rij bevat de itemparameters voor toetsversie SZC. Voor dit item is er niet veel verschil in de discriminatie waarde voor de toetsversies, maar de moeilijkheidsparameter voor SZC is veel lager.

Tabel 6.7a Schatting IRT itemparameters toetstaak Herkennen

ITEM	FMC	FZC	SMC	SZC	a_i (discriminatie parameter)	b_i (moeilijkheids parameter)	Informatie waarde
1	1	1	1	1	0,546	-2,034	0,056
3	1	1	1	1	0,765	-1,569	0,104
9	1	1	1	1	0,258	-1,406	0,016
10	1	1	1	1	0,372	-0,472	0,034
14	1	1	1	1	0,764	-2,806	0,055
15	1	1	1	1	0,886	-2,621	0,064
16	1	1	1	0	0,785	-1,570	0,108
16	0	0	0	1	0,858	-3,645	0,030
17	1	1	1	1	0,413	-2,138	0,035
18	1	1	1	1	1,306	-2,511	0,060
19	1	1	1	1	1,083	-1,822	0,126
20	1	1	1	1	0,855	-2,879	0,053
25	1	1	1	1	0,613	0,500	0,092
31	1	1	1	1	0,295	0,007	0,022
36	1	1	1	1	0,251	-1,834	0,015
37	1	1	1	1	0,922	-2,239	0,085
41	1	1	1	1	0,500	-0,966	0,059
49	1	1	1	0	0,379	-0,601	0,035
49	0	0	0	1	3,184	-3,346	0,000
57	1	1	1	1	0,251	-1,834	0,015
58	1	1	1	1	0,337	-2,536	0,024

Met de twee itemparameters a_i en b_i is de bijdrage aan de betrouwbaarheid van de toets voor een specifieke student te berekenen. De laatste kolom van de drie tabellen geeft de zogenaamde informatiewaarde van het item, dat is de bijdrage aan de betrouwbaarheid van dat item aan de toets voor een specifieke leerling. De informatiewaarden zijn hier uitgerekend voor een gemiddelde leerling. De informatiewaarde verschilt over leerlingen: een item dat veel bijdraagt voor leerlingen met een hoge vaardigheid draagt vaak weinig bij voor leerlingen met een lage vaardigheid. Items met hoge discriminatie parameters dragen meer bij dan items met een lage discriminatieparameter en informatie is maximaal voor een vaardigheid gelijk aan de moeilijkheidsparameter. Deze eigenschappen van informatie zijn van nut bij het selecteren van items voor specifieke doelgroepen leerlingen. Ze vormen ook de basis voor computer-adaptief toetsen waarbij items worden gegeven die het best passen bij de vaardigheid van de specifieke leerling.

Merk hierbij op dat item 49 een in toetsversie SZC een extreem item is: het heeft een discriminatieparameter van 3,184 en een moeilijkheidsparameter van -3,346.

Dit wordt bevestigd in de toetsitem analyse in tabel 6.1c: daar heeft het item een gemiddelde van 0,89 en een punt-biseriële correlatie van 0,476. De punt-biseriële correlatie lijkt te wijzen op een hoge bijdrage aan de betrouwbaarheid maar de IRT analyse leidt tot een andere conclusie zoals uit de laatste kolom van tabel 6.7a blijkt.

Tabel 6.7b *Schatting IRT itemparameters Toetstaak Benoemen*

ITEM	FMC	FZC	SMC	SZC	a_i (discriminatie parameter)	b_i (moeilijkheids parameter)	Informatie waarde
2	0	1	1	1	0,436	-0,003	0,048
2	1	0	0	0	0,250	2,370	0,014
8	1	1	1	1	0,604	-1,427	0,076
13	1	1	1	1	0,679	-0,951	0,104
21	1	1	1	1	1,381	-1,874	0,124
23	1	1	1	1	1,145	-1,384	0,185
24	1	1	1	1	1,449	-1,858	0,125
26	1	1	1	1	1,541	-0,955	0,361
27	0	0	1	0	1,498	-2,232	0,074
27	0	1	0	0	0,308	-0,079	0,024
27	1	0	0	0	2,391	-0,616	0,867
27	0	0	0	1	0,264	-2,357	0,016
29	1	1	1	1	1,002	-0,858	0,210
30	1	1	1	0	1,109	-0,608	0,275
30	0	0	0	1	1,972	0,565	0,723
32	1	1	1	1	0,975	-1,091	0,181
33	1	1	1	1	0,448	-1,374	0,046
38	1	1	0	1	1,242	-1,653	0,156
38	0	0	1	0	0,429	0,002	0,046
39	1	1	1	1	1,318	-2,337	0,073
40	1	1	1	1	1,489	-0,588	0,460
42	1	1	0	1	1,078	-0,935	0,228
42	0	0	1	0	1,298	-2,444	0,065
43	1	1	1	1	1,972	-1,886	0,090
44	1	0	1	1	2,256	-2,206	0,035
44	0	1	0	0	3,889	-5,311	0,000
45	1	1	1	1	2,316	-2,097	0,041
56	0	1	1	0	0,858	-2,961	0,050
56	1	0	0	0	1,268	-1,537	0,175
56	0	0	0	1	0,858	3,904	0,024

Tabel 6.7c *Schatting IRT itemparameters Toetstaak Relaties*

ITEM	FMC	FZC	SMC	SZC	a _i (discriminatie parameter)	b _i (moeilijkheids parameter)	Informatie waarde
4	1	1	1	1	1,013	-2,394	0,077
5	1	1	1	1	0,654	-1,899	0,074
6	1	1	1	1	0,326	0,208	0,027
7	1	1	1	1	0,353	0,801	0,031
11	1	1	1	1	0,466	1,337	0,049
12	1	1	1	1	0,916	0,654	0,192
22	1	1	1	1	0,794	0,353	0,155
34	1	1	1	1	0,718	-1,656	0,092
35	1	1	1	1	0,769	-0,023	0,148
46	1	1	0	0	0,886	0,450	0,189
46	0	0	0	1	3,552	-2,616	0,001
46	0	0	1	0	1,069	-1,499	0,159
47	1	1	1	1	1,274	0,046	0,405
48	1	1	0	0	1,133	0,420	0,303
48	0	0	0	1	2,754	-2,231	0,016
48	0	0	1	0	0,899	-1,218	0,152
50	1	1	1	1	1,005	-0,119	0,252
51	1	1	1	1	0,658	-0,578	0,104
52	1	1	1	1	1,093	-1,868	0,121
53	1	1	1	1	2,006	-0,779	0,576
54	1	0	1	1	0,979	-0,501	0,226
54	0	1	0	0	0,615	1,937	0,068
55	1	1	1	1	0,858	2,983	0,049

Het aantal toegevoegde itemparameters is relatief beperkt: het proces van het zoeken van een passend model begon met de 57 discriminatie en 57 moeilijkheidsparameters, en eindigde met 57+17=74 van die twee soorten parameters. Over het algemeen hebben de verschillende toetsversies dus dezelfde itemparameters. Ook hieruit kunnen we concluderen dat de drie deelconstructen in de verschillende toetsversies vergelijkbaar verschillend functioneren.

Het eindresultaat is dat:

1. De vier toetsversies globaal dezelfde constructen meten omdat
  - a. De patronen in de correlatiematrices goed vergelijkbaar zijn;
  - b. Het merendeel van de items in de vier toetsversies hetzelfde functioneert.
2. Er wel verschillen zijn:
  - a. Alleen de correlatiematrices van FMC en FZC verschillen niet significant.
  - b. 17 van de 57 items functioneren verschillend over de toetsversies.

3. Het goed mogelijk blijkt om een gezamenlijk model voor de hele dataverzameling te creëren. Daarom zijn de toetsen goed te equivaleren: een score op de ene toets is vertaalbaar naar een vergelijkbare score op een andere toets.

Afwijkende en zwakke items nader bekeken

Met de analyses die met de klassieke testtheorie (tabellen 6.1a tot en met 6.1d) zijn uitgevoerd zijn items geïdentificeerd die een lage variantie of punt-biseriële correlatie hebben en die significant verschillen tussen de toetsversies. Met de analyses die met de itemresponstheorie (tabellen 6.5 tot en met 6.7c) zijn uitgevoerd zijn items geïdentificeerd die zich verschillend gedragen in de toetsversies daar hun item parameters over deze versies heen verschillen.

In onderstaande tabel 6.8 zijn deze items weergegeven.

Tabel 6.8 Overzicht van afwijkende en zwakke items

Itemnr	Toets-taak	Zwak in toetsversie	Significant verschil T-toetsen	Afwijkend in IRT-model
1	H	SMC		
3	H	FZC		
9	H	FMC, FZC, SZC		
10	H	FMC, FZC, SZC		
14	H	FZC, SMC, SZC		
15	H	FMC, SZC		
16	H	SZC	SZC t.o.v. FMC, FZC, SMC	
17	H	FMC, FZC, SMC, SZC	SMC t.o.v. FMC	
18	H	SMC, SZC	SMC, SZC t.o.v. FZC	
19	H	SMC,		
20	H	FMC, SMC, SZC	SMC t.o.v. FZC	
31	H	SZC		
36	H	FMC, SMC, SZC		
37	H	SZC		
41	H	SMC, SZC	SMC t.o.v. FMC, FZC SZC t.o.v. FZC	
49	H	SMC	SZC t.o.v. FMC, FZC	SZC
57	H	FZC, SMC, SZC		
58	H	FMC, FZC, SMC, SZC		
2	B	FZC, SMC	SMC t.o.v. FMC, FZC, SZC	
8	B	FMC		
21	B		SMC t.o.v. FMC	

Tabel 6.8 *Overzicht van afwijkende en zwakke items*

Itemnr	Toets- taak	Zwak in toetsversie	Significant verschil T- toetsen	Afwijkend in IRT- model
23	B		SMC t.o.v. FMC	
26	B		SMC t.o.v. FMC, FZC	
27	B	SZC	SMC, SZC t.o.v. FMC, FZC	FMC, FZC, SMC, SZC
30	B		SMC t.o.v. SZC	SZC
32	B	SZC	SMC t.o.v. SZC	
33	B	FMC, FZC, SZC		
38	B		SMC t.o.v. FMC, SZC	SMC
39	B	SZC		
42	B		SMC t.o.v. FMC	SMC
43	B		SMC t.o.v. FMC	
44	B	SZC		FZC
56	B	SMC, SZC	FMC, FZC t.o.v. SMC, SZC	FMC, SZC
4	R	FZC, SZC	SZC t.o.v. FMC, SMC	
5	R	FZC, SMC, SZC		
6	R	SZC		
7	R	FMC, FZC, SZC		
11	R	SMC, SZC		
12	R	SZC		
22	R	SZC		
34	R	SZC		
35	R	SZC		
46	R		FMC, FZC t.o.v. SMC, SZC	SMC, SZC
47	R		SZC t.o.v. FMC, FZC	
48	R		FMC, FZC t.o.v. SMC, SZC	SMC, SZC
50	R	FMC		
51	R	FZC		
52	R	FZC		
54	R	FZC	FZC t.o.v. FMC, SMC, SZC	FZC
55	R	FMC, FZC, SMC, SZC		

In de tabel is in de derde kolom te zien welke items zwak zijn in een bepaalde toetsversie waarbij alleen items 17, 55 en 58 zwak zijn in elke toetsversie. Deze items dienen bij verder gebruik te worden verbeterd. In kolom 4 zijn de items weergegeven waarvan met T-testen is aangetoond dat er significante verschillen zijn in gemiddelde score op het betreffende item tussen de toetsversies. In de ene toetsversie is het item dus veel gemakkelijker of moeilijker in vergelijking tot

datzelfde item in een andere toetsversie. Dit wordt waarschijnlijk veroorzaakt door de representatievorm van dat item. In kolom 5 zijn de items weergegeven die afwijkend functioneren in de IRT-modellering en op één item na, item 44, komen deze items overeen met de items die significant verschillen tussen enkele toetsversies volgens de T-toetsen, kolom 4. Ook hier is het plausibel dat de representatievorm hier een rol in speelt. Beide analyses gecombineerd bieden robuustere bewijsvoering van items die mogelijk verschillend functioneren afhankelijk van de representatievorm. Dat zijn de items 2, 16, 27, 30, 38, 42, 46, 48, 49, 54 en 56. Tabel 6.9 geeft een overzicht van de visualisaties in deze items. Item 44, dat weliswaar afwijkend functioneert maar niet significant verschilt tussen de representatievormen, is ook meegenomen. Deze items zouden nader bestudeerd kunnen worden om te bezien of op grond van de relatie tussen de visualisatie, de inhoud van de plaatjes en de toetstaak zou kunnen worden begrepen waarom deze items verschillend functioneren in de verschillende toetsversies. Op deze plaats is dat echter te vroeg. Voortgezette analyse van deze items bij nieuwe toetsafnames, zou eerst moeten aantonen dat de gevonden effecten zich herhalen.

Tabel 6.9 Visualisatievormen van de in tabel 6.8 geïdentificeerde items met afwijkend gedrag

Toetsversie		FMC	FZC	SMC	SZC
Vraag	Taak	Vraag			
Thema 1: Buikholte en spijsvertering					
Q02	B	Benoem het aangegeven onderdeel van de rattemaag. (ventraal aanzicht)			
Thema 2: Spieren					
Q16	H	In de afbeelding worden de verschillende spieren met een pijl aangegeven. Welk deel van de spieren is een weergave van de musculus rectus abdominus ? (ventraal aanzicht)			
Thema 3: Geslachtsorganen					
Q27	B	Benoem het aangegeven anatomisch onderdeel. E= deel epididymis, *= ductuli efferentes (ventraal aanzicht)			
Q30	B	Benoem de aangegeven klier van de rechter mannelijke geslachtsorganen. (ventraal aanzicht)			

Tabel 6.9 Visualisatievormen van de in tabel 6.8 geïdentificeerde items met afwijkend gedrag

Toetsversie		FMC	FZC	SMC	SZC
Vraag	Taak	Vraag			
Thema 4: Borstholte en grote bloedvaten					
Q38	B	Benoem het aangegeven anatomisch onderdeel. (ventraal aanzicht)			
Q42	B	Benoem het aangegeven anatomisch onderdeel. (ventraal aanzicht)			
Q44*	B	Benoem het aangegeven anatomisch onderdeel. (ventraal aanzicht)			
Q46	R	Geef aan op welk deel van het hart de grote lichaamsvenen zijn aangesloten. VA= vena azygos, VCC= vena cava caudalis, *= aansluiting bloedvat op het hart. (ventraal aanzicht)			
Q48	R	Geef aan op welk deel van het hart de aortaboog is aangesloten. AC= arteria carotis, AD= aorta dorsalis, TB= truncus brachiocephalicus, *= aansluiting bloedvat op het hart. (ventraal aanzicht)			
Q49	H	Welk plaatje is een correcte weergave van het hart met aan- en afvoerende bloedvaten? (ventraal aanzicht)			
Thema 5: (Bij)nieren en bloedvaten					
Q54	R	Vanuit ventraal aanzicht zijn de lichaamsslagader en de onderste holle ader weergegeven. Verbind op de juiste wijze – denk aan de bloedstroomrichting – deze bloedvaten met de rechter nier en met de rechter achterpoot.			
Q56	B	In het onderstaand plaatje wordt de stroomrichting van het bloed in de arteriën en venen aangegeven (rode pijlen). Benoem de met een ster aangegeven arteriën en venen. (ventraal aanzicht)			

Protocolanalyse

Voor de protocol-analyse zijn de antwoorden van de studenten geanalyseerd van zowel de mondelinge variant als de schriftelijke variant. Van de mondelinge variant is een aantal geluidsopnamen afgeluisterd en een transcriptie gemaakt waarbij is geconstateerd dat de inhoud van de responsen geen toegevoegde waarde heeft aan wat de schriftelijke variant oplevert. In combinatie met de tijdrovende analyse vereist bij de mondelinge variant is de meeropbrengst gering met als gevolg dat is afgezien van verdere verwerking van deze data.

Het databestand van de schriftelijke protocolanalyse telde 2655 antwoordposities. De studenten blijken soms vragen te hebben overgeslagen. In de analyse zijn alleen niet-lege responsen meegenomen. Dat leverde 2501 bruikbare responsen op. Dat is voor de 57 toetsitems gemiddeld 43,9 responsies per item. De resultaten zijn geanalyseerd met behulp van trefwoorden. Bijlage 6.1 geeft een overzicht van de gebruikte trefwoordcombinaties.

Met de gebruikte methode kon 77,2 % van de responsen worden geclassificeerd naar aanleiding van de trefwoorden waarna de trefwoorden werden geclassificeerd en totalen berekend. De overige, niet geclassificeerde, responsen bevatten mededelingen als “wist ik niet”, “wist ik gewoon”, “zelfde als vorige vraag”, of soms alleen “?”. In figuur 6.4 staan enkele voorbeelden van responsen met bijbehorende classificaties. In bijlage 6.2 is een overzicht van de berekende totalen van de classificaties gepresenteerd.

Soort respons	Voorbeelden
Verwijzing naar duidelijkheid of onduidelijkheid plaatje	"ik weet omdat ik het net heb geleerd en omdat het duidelijk op de poster was te zien, "; "alleen op de foto is het onduidelijk"
Verwijzing naar de vormkenmerken van anatomische objecten	"vorm object"; " grote gelobde klier een rechte doorgang had naar de ductus urogenitali"; "naar de vorm ervan gekeken"
Verwijzing naar locatie en oriëntatie van anatomische objecten	"orientatie spier"; " de richting loopt van midden hoog naar buiten"; "de gang naar de zaadbal komt onder in de zaadbal uit";
Verwijzing naar instructiecomponenten	"door herinnering van dissectie die ik zelf heb uitgevoerd"; "Onthouden door practica tijdens het voorsnijden"; "net geleerd uit syllabus"; "door onze fantastische student-assistent, die dit ontzettend goed heeft uitgelegd"; "bij de nabespreking opgestoken".
Uitleg antwoord vanuit de leerstof	"de duodenum is het eerst orgaan dat uit de maag komt. dit weet in naar aanleiding van een van de herkenningpunten van de maag. duodenum (=twaalfvingerige darm) is dan het eerste orgaan."; "Het coecum is een soort zak. De dikke en dunne darm zitten aan dezelfde kant aan het coecum vast en het coecum hangt er als het ware tussen in. Dit is niet zo in de andere coecum waar het voedsel door het coecum heen moet om bij de dikke darm te komen."

Figuur 6.4 Voorbeelden van geclassificeerde responsen.

Kwaliteit instructie-toetscombinatie

In tabel 6.10 is voor wat betreft de aansluiting instructie en toets te zien dat er verwijzingen zijn naar alle instructiecomponenten. De percentages in de laatste kolom laten zien dat soms ook wordt verwezen naar een combinaties van instructiecomponenten. Het meeste (66%) wordt verwezen naar de dissectie zelf en/of het hebben gezien van het object in het practicum en/of de tekening die werd aangeboden tijdens elk dissectieonderdeel. De begeleiding is een goede tweede met 27%, de nabespreking wordt in 13% van de gevallen genoemd. In de laatste kolom van Tabel 6.10 valt op dat het meest verwezen wordt naar instructiecomponenten tijdens de herkenningssitens en dat dit afneemt voor benoemen en vervolgens ook voor relaties. Met een chi-kwadraat toets (χ^2 -toets) naar de afhankelijkheid van instructiecomponenten en de toetstaken is nagegaan of deze relatie inderdaad aanwezig is. Uit het resultaat onderaan de tabel blijkt dat deze twee variabelen onafhankelijk zijn en er geen relatie is.

Tabel 6.10 Totalen voor verwijzingen naar de instructiecomponenten

	Totaal syllabus / handleiding	Totaal voorsnijden / voorbespreking	Totaal dissectie / gezien in practicum /tekening	Totaal nabespreking	Totaal begeleiding	Responzen met verwijzing naar instructiecomponent
Totalen voor herkennen:	23	15	242	51	106	351
Als percentage van responzen met verwijzing naar instructiecomponent:	7	4	69	15	30	125
Totalen voor benoemen:	37	15	199	44	73	313
Als percentage van responzen met verwijzing naar instructiecomponent:	12	5	64	14	23	118
Totalen voor relaties:	19	16	168	26	73	257
Als percentage van responzen met verwijzing naar instructiecomponent:	7	6	65	10	28	116
Algehele totalen:	79	46	609	121	252	921
Als percentage van responzen met verwijzing naar instructiecomponent:	9	5	66	13	27	120

Noot: $X^2 = 13,37$; $df = 8$; $p = 0,06$. De percentages in de laatste kolom zijn de optellingen van de andere percentages in de betreffende rij. Omdat soms in een respons meer dan één instructiecomponent wordt genoemd, zijn deze percentages groter dan 100%.

Het algemene beeld van tabel 6.10 is dat de combinatie van instructiecomponenten van de vernieuwde instructie goed functioneert.

In tabel 6.11 zijn de aantallen van verwijzingen naar instructiecomponenten weergegeven evenals de aantallen van inhoudelijke beschrijvingen van de leerstof met daaraan toegevoegd de percentages. Van het totaal aantal antwoorden is te zien dat voor herkennen en benoemen percentueel vaker wordt verwezen naar de instructiecomponenten als dat er inhoudelijke beschrijvingen worden gegeven. Voor de toetstaak relaties wordt veel meer gebruik gemaakt van een inhoudelijke beschrijving en wordt naar verhouding veel minder verwezen naar de instructieonderdelen. Men beschrijft hoe men tot het antwoord is gekomen met elementen uit de leerstof zoals de objecten, de relaties tussen de objecten en de functies. Met een chikwadraat toets is nagegaan of deze relatie inderdaad aanwezig is en uit het resultaat blijkt dat dit het geval is. Dit spoort met de eerdere constatering dat de drie taken herkennen, benoemen en relaties in die

volgorde een stijgende moeilijkheidsgraad kennen. Voor de relatietaak moet meer inhoudelijk worden beredeneerd dan voor de andere taken.

Tabel 6.11 *Instructie versus inhoud*

	Totaal responsen met verwijzing naar instructiecomponenten	in % van het totaal aantal responsen	Beschrijvend met elementen van de leerstof	in % van het totaal aantal responsen	Totaal aantal niet-lege responsen
Herkennen	351	42	250	30	841
Benoemen	313	35	205	23	887
Relaties	257	33	325	42	773
Totaal	921	37	780	31	2501

Noot: $\chi^2 = 57,1$; $df = 2$; $p = 0,00$.

Invloed representatievorm

In tabel 6.12 zijn de aantallen weergegeven van verwijzingen naar de kenmerken kleur, locatie, oriëntatie en vorm van een object. In de achtste kolom staat hoe vaak in een respons naar een of meer van de vormkenmerken wordt verwezen. In de negende kolom is meegenomen hoe vaak een plaatje of foto onduidelijk wordt gevonden. Uit het resultaat van de onderaan de tabel vermelde chi-kwadraattoets blijkt dat er significante interacties zijn tussen de toetstaken en de verwijzing naar de objectkenmerken. In de responsen wordt met regelmaat gerefereerd naar kenmerken zoals locatie, oriëntatie en vorm en soms ook naar kleur. Refereren aan kleur was overigens alleen mogelijk voor de helft van de studenten die de toetsversies met foto's kregen. De vormkenmerken lijken te worden gezien als kenmerken van de leerstof waarmee de student direct interacteert. Over de plaatjes of foto's wordt alleen iets gezegd als ze onduidelijk worden gevonden. De aantallen daarvoor zijn klein in relatie tot het totaal aantal responsen. Uit dat oogpunt lijken de visualisaties te voldoen.

Tabel 6.12 *Inloed representatievorm*

	Totaal kleuren	Kleuren %	Totaal locatie / oriëntatie	Locatie/ oriëntatie %	Totaal vorm	Vorm %	Totaal antwoorden met verwijzing naar vormkenmerken	Verwijzing naar onduidelijke plaatjes/ foto	Verwijzing naar onduidelijke plaatjes/ foto %	Totaal aantal niet-lege responsen
Herkennen	14	1,7	214	25	72	8,6	281	5	0,6	841
Benoemen	17	1,9	128	14,5	62	7	182	1	0,1	887
Relaties	4	0,6	132	17	33	4,3	162	5	0,6	773
Totaal	35	1,4	474	19	167	6,7	625	11	0,4	2501

Noot: $X^2 = 50,81$; $df = 6$, $p = 0,00$.

6.4 DISCUSSIE EN CONCLUSIE

Het hoofddoel van het hier beschreven experiment is de evaluatie van de ontwikkelde instructie-toetscombinatie waarbij tevens wordt geprobeerd uitsluitsel te krijgen over de vraag welke vormgeving van de visuele representaties in de toets de beste keuze is met betrekking tot de factoren realisme (schema's of foto's) en context (wel of geen). Eerst wordt ingegaan op de instructie-toetscombinatie en vervolgens worden de onderzoeksvragen besproken.

Instructie-toetscombinatie

De kwaliteit van de toets

De toets zoals gepresenteerd in dit hoofdstuk, een tweede versie van de ontwikkelde CBT-toets, is als geheel op basis van de betrouwbaarheidswaarden verbeterd in vergelijking tot het vorige ontwerp met uitzondering van SZC. Voor elke toetsversie, als geheel, is de waarde van Cronbachs alpha in vergelijking tot de vorige ongeveer gelijk gebleven en wordt ruim voldaan aan de minimale ondergrens. De toets is dus in elke representatievorm stabiel in vergelijking tot de eerdere versie.

Voor het toetsonderdeel herkennen is een verbetering te zien ten opzichte van de vorige versie voor de toetsversies met foto's die nu wel net voldoen aan de

minimale betrouwbaarheidsgrens. Voor de toetsversies met schema's is de kwaliteit verminderd, de alpha-waarden zijn laag en ver onder de minimale betrouwbaarheidsgrens. Dit toetsonderdeel blijft ook in de tweede ontworpen versie een zwak onderdeel.

De kwaliteit van het toetsonderdeel benoemen is lichtelijk verbeterd in vergelijking tot de eerdere ontworpen versie. De toetsversies voldoen elk aan de minimale betrouwbaarheidsgrens met min of meer vergelijkbare betrouwbaarheid als de eerdere ontworpen versie voor de toetsversies foto met context en de schematische toetsversies.

Voor het toetsonderdeel relaties is een lichte kwaliteitsverbetering te zien voor de toetsversies met context. Deze toetsversies voldoen aan de betrouwbaarheidsgrens waarbij de versie foto met context kwalitatief is verbeterd. De toetsversies zonder context zijn in vergelijking tot de eerdere ontworpen versies minder betrouwbaar.

Voorts blijkt dat de gemiddelde toetsscore per toetstaak afneemt van herkennen naar benoemen naar relaties met significante verschillen duidend op een moeilijkheidsgraad. Dit lijkt niet heel verwonderlijk: mogelijk is een voorwaarde om te kunnen benoemen het herkennen van de betreffende anatomische structuren en om relaties te kunnen leggen moeten de betrokken anatomische structuren kunnen worden geïdentificeerd. Mogelijk is er tussen de toetstaken een hiërarchie in de lijn van de klassieke literatuur over leerhiërarchieën (Gagné, 1977; Bloom et al, 1994; Jonassen, Tessmer & Hannum, 1999). Zoals in hoofdstuk 2 is vermeld ontwierp Dwyer (1978) op basis van Gagné's theorie over leercondities zowel de instructie als de toets van het PSE-programma. Zo maakt ook Hegarty (2005) een onderscheid in kennistypen maar geeft zij vervolgens geen uitsluitel over een mogelijke hiërarchische ordening. Vergelijkbaar geven Martinez en Katz (1992) ook aan dat er een verschil is tussen CFR-itemformats waarbij 'discovered strategy items' het moeilijkst zijn. Ook zij doen geen expliciete uitspraken over een hiërarchische ordening.

Analyse op itemniveau wijst op een kwalitatieve verbetering van de toets in de verschillende toetsversies omdat er minder zwakke items zijn in vergelijking tot de vorige ontworpen versie. De zwakke items zijn vooral in de toetsversies met schema's en hebben veelal een hogere gemiddelde waarde dan de versies met

foto's. Dat verschil is voor een aantal items ook significant hetgeen duidt op een verschil in moeilijkheid voor een specifiek item in de verschillende representatievormen.

Voor wat betreft de aansluiting instructie en toets lijkt deze voor studenten adequaat daar bij navraag over de wijze van invullen van de toets de studenten overwegend instructiecomponenten aanhalen als herinnering of een beschrijving geven van de leerstof voor het beantwoorden van de vragen. Naar representatievormen wordt weinig gerefereerd.

Vergelijkbaarheid van de constructen van de toetsversies

De vergelijkbaarheid tussen de constructen in de verschillende toetsversies is onderzocht door na te gaan of de relaties tussen de drie deelconstructen en de eigenschappen van de items in alle vier de toetsversies vergelijkbaar zijn. De eindconclusie uit de analyses is dat de vier toetsversies globaal dezelfde constructen meten omdat de patronen in de correlatiematrices goed vergelijkbaar zijn en het merendeel van de items in de vier toetsversies hetzelfde functioneert. Hoewel vergelijkbaar zijn de toetsversies niet hetzelfde. Ten behoeve van de analyse zijn alle vier versies met elkaar vergeleken waarbij voor de analyse alle items zijn samengevoegd. Er zijn dus 57 combinaties van de vier versies vergeleken met als resultaat dat in 12 itemcombinaties in totaal 17 items verschillend functioneren over de toetsversies. Over het geheel genomen blijkt het goed mogelijk een gezamenlijk model voor de dataverzameling te creëren.

Invloed van de factoren realisme en context

Eén van de wijzigingen ter verbetering van de toets was de verduidelijking van de factor context door meer onderscheid aan te brengen tussen de aan- of afwezigheid van context in een representatievorm. Deze wijziging heeft geen effect gehad omdat er geen significante verschillen zijn gevonden voor de factor context. Ook deden zich geen interactie-effecten voor tussen realisme en context. Wel is er een significant effect voor de factor realisme voor zowel het herkennen van anatomische structuren als het leggen van functionele relaties in het voordeel van schema's. Dit resultaat is precair voor de toetstaak herkennen aangezien de gemiddelden van schematische toetsen weliswaar hoger zijn dan voor foto's maar de betrouwbaarheidswaarden voor de toetsversies met schema's lager zijn en ver onder de betrouwbaarheidsnorm liggen. Voor de toetstaak relaties geldt hetzelfde met het verschil dat de toetsversie schema met context wel voldoet aan de betrouwbaarheidsnorm.

Voor het achterhalen van mogelijke verschillen op itemniveau is met behulp van tabel 6.8 een aantal verschillend functionerende items geselecteerd waarvan de resultaten uit de IRT-analyse en van de T-toetsen elkaar versterken. Mogelijk is het verschillend gedrag van deze items in de verschillende toetsversies gelegen in de verschillen in representatievorm. Zoals al opgemerkt kan daarover pas meer uitsluitend worden verkregen als deze items bij nieuwe toetsafnames worden hergebruikt om te zien of het geconstateerde gedrag van deze items persisteert.

Resumerend kan, ten aanzien van de instructie-toetscombinatie worden geconcludeerd dat de toets als geheel verbeterd is ten opzichte van de eerste CBT-versie met uitzondering van de versie schema zonder context. De toets is betrouwbaarder en er is sprake van meer onderlinge samenhang tussen de items. Het onderdeel herkennen blijkt het stabielst voor de toetsversies met foto's. Het toetsonderdeel benoemen blijkt een stabiel onderdeel voor elke toetsversie. Het toetsonderdeel relaties blijkt het stabielst voor de toetsversies met context met name de versie schema met context.

Voorts blijkt de kwaliteit van de toets voldoende voor gebruik waarbij is aangetoond dat de vier toetsversies globaal dezelfde constructen meten en als zodanig goed zijn te equivaleren. De mogelijkheid bestaat hierdoor om op basis van deze items nieuwe toetsen samen te stellen waarin gevarieerd wordt met de representatievormen. De tabellen 6.13 a tot en met c tonen per toetstaak in welke representatievormen een specifiek item bruikbaar is.

Tabel 6.13a *Herkennen, bruikbare itemvarianten*

Items	FMC	FZC	SMC	SZC	Aantal bruikbare varianten
1	✓	✓	✓	✓	4
3	✓	✓	✓	✓	4
9	✓	✓	✓	✓	4
10	✓	✓	✓	✓	4
14	✓	zwak	zwak	zwak	1
15	zwak	✓	✓	zwak	2
16	✓	✓	✓	zwak	3
17	zwak	zwak	zwak	zwak	Geen
18	✓	✓	zwak	zwak	2
19	✓	✓	zwak	✓	3
20	zwak	✓	zwak	zwak	1
25	✓	✓	✓	✓	4

Tabel 6.13a *Herkennen, bruikbare itemvarianten*

Items	FMC	FZC	SMC	SZC	Aantal bruikbare varianten	
31	✓	✓	✓	✓	4	
36	zwak	✓	✓	✓	3	
37	✓	✓	✓	zwak	3	
41	✓	✓	✓	✓	4	
49	✓	✓	✓	✓	4	
57	✓	✓	✓	zwak	3	
58	zwak	zwak	zwak	zwak	Geen	
	FMC	FZC	SMC	SZC	Totaal	
aantal bruikbaar	14	16	13	10	53	70%
aantal zwak	5	3	6	9	23	30%
Totaal	19	19	19	19	76	100%
✓ = bruikbaar item						

Tabel 6.13b *Benoemen, bruikbare itemvarianten*

Items	FMC	FZC	SMC	SZC	Aantal bruikbare varianten	
2	✓	✓	zwak	✓	3	
8	✓	✓	✓	✓	4	
13	✓	✓	✓	✓	4	
21	✓	✓	✓	✓	4	
23	✓	✓	✓	✓	4	
24	✓	✓	✓	✓	4	
26	✓	✓	✓	✓	4	
27	✓	✓	✓	zwak	3	
29	✓	✓	✓	✓	4	
30	✓	✓	✓	✓	4	
32	✓	✓	✓	✓	4	
33	✓	✓	✓	✓	4	
38	✓	✓	✓	✓	4	
39	✓	✓	✓	zwak	3	
40	✓	✓	✓	✓	4	
42	✓	✓	✓	✓	4	
43	✓	✓	✓	✓	4	
44	✓	✓	✓	zwak	3	
45	✓	✓	✓	✓	4	
56	✓	✓	zwak	zwak	2	
	FMC	FZC	SMC	SZC	Totaal	

Tabel 6.13b *Benoemen, bruikbare itemvarianten*

Items	FMC	FZC	SMC	SZC		Aantal bruikbare varianten
aantal bruikbaar	20	20	18	16	74	93%
aantal zwak	0	0	2	4	6	7%
Totaal	20	20	20	20	80	100%
✓ = bruikbaar item						

Tabel 6.13c: *Relaties, bruikbare itemvarianten*

Items	FMC	FZC	SMC	SZC		Aantal bruikbare varianten
4	✓	zwak	✓	zwak		2
5	✓	zwak	✓	✓		3
6	✓	✓	✓	✓		4
7	✓	✓	✓	✓		4
11	✓	✓	✓	✓		4
12	✓	✓	✓	✓		4
22	✓	✓	✓	✓		4
34	✓	✓	✓	✓		4
35	✓	✓	✓	✓		4
46	✓	✓	✓	✓		4
47	✓	✓	✓	✓		4
48	✓	✓	✓	✓		4
50	✓	✓	✓	✓		4
51	✓	✓	✓	✓		4
52	✓	✓	✓	✓		4
53	✓	✓	✓	✓		4
54	✓	✓	✓	✓		4
55	zwak	zwak	zwak	zwak		Geen
	FMC	FZC	SMC	SZC	Totaal	
aantal bruikbaar	17	15	17	16	65	90%
aantal zwak	1	3	1	2	7	10%
Totaal	18	18	18	18	72	100%
✓ = bruikbaar item						

Met bovenstaande tabellen is de gebruiker, in dit geval de docentenstaf van de vakgroep dierfysiologie, in staat om dezelfde toets met meerdere representatievormen aan te bieden aan de studenten daar veel items geschikt zijn voor verder gebruik. Alvorens dit mogelijk is dient de docentenstaf nog keuzes te maken in de samenstelling en doorontwikkeling van de toets. De huidige

programmatuur vereist ook actualisatie op technisch vlak alvorens deze kan worden geïmplementeerd.

Voor wat betreft de samenstelling van de toets kan de docentenstaf kiezen voor een samengestelde toets waarin een gebalanceerde mix van representatievormen of kiezen voor een toets met één specifieke representatievorm. In het laatste geval kunnen daarbij in de meeste gevallen zwakke items worden omzeild door een bruikbaar item te kiezen uit de meest verwante toetsvorm (bij de keuze voor FMC bijvoorbeeld in plaats van het zwakke item 15 het overeenkomstige item uit FZC, enzovoort). Een mix van representatievormen is mogelijk omdat elk item immers, ongeacht de in dit onderzoek toegepaste visualisatievorm, hetzelfde construct meet. De keuze voor een visualisatievorm(en) is vooral gelegen in het belang dat de docentenstaf daaraan hecht gegeven de leersituatie waarin met een reëel object wordt geleerd. Wel dient de staf zich ervan bewust te zijn dat items met schema's zeer waarschijnlijk gemakkelijker zijn dan items met foto's en is het noodzakelijk het gedrag van de items te volgen. In geval wordt gekozen voor een samengestelde toets met een mix van representatievormen betekent dit de ontwikkeling van een nieuwe toets die bij evaluatie nieuwe informatie kan opleveren over het gedrag van de items in relatie tot de representatievormen. Dit vereist dat bij elke toetsafname de toets wordt geanalyseerd om meer te weten over de items op basis van de gegenereerde data tot blijkt dat er een stabiel effect zichtbaar is in het gedrag van de items. Ook is daarbij van belang een analysemethodiek te hanteren waarmee wordt nagegaan hoe visualiteit zich verhoudt tot het meten van het construct.

HOOFDSTUK 7

Visueel toetsen van Visueel leren

In dit afsluitend hoofdstuk wordt de balans opgemaakt voor wat betreft de kwaliteit van de instructie-toetscombinatie evenals de bijdrage aan kennis over visueel leren (de instructie) en visueel toetsen (de toets) met betrekking tot de variabelen realisme en context.

7.1 ONDERZOEKSPROBLEMATIEK EN ONDERZOEKSAANPAK

De directe aanleiding voor het hier gepresenteerde onderzoek is het in hoofdstuk 1 geschetst toetsprobleem voor dissectiepractica zoals deze van oudsher worden geïnstreerd en getoetst aan de faculteit Biologie aan de Rijksuniversiteit Groningen. De huidige faculteitsnaam is faculteit Wiskunde en Natuurwetenschappen. Toetsing vindt daarbij plaats met een beoordeling van de tijdens de dissectie door een student geproduceerde set tekeningen. Een toetsingsmethodiek die om twee redenen niet wenselijk is. De beoordeling van tekeningen is een tijdrovend en arbeidsintensief proces waarbij studenten verlaat feedback ontvangen op hun leerproces. De objectiviteit van deze toetsmethodiek is niet gegarandeerd omdat de beoordeling verdeeld is over meerdere personen (de docent en studentassistenten), er onzekerheid is over de beoordelingscriteria en de beoordelaarssamenstelling verschilt per practicum met zowel ervaren als onervaren beoordelaars. Dit onderzoek is gestart om deze problemen het hoofd te bieden met als opdracht een toets te ontwikkelen die (a) aansluit op het leerproces, (b) objectief is, (c) studenten in een relatief kort tijdsbestek van feedback voorziet en (d) een alternatief biedt voor het handmatig corrigeren van tekeningen.

Uit een eerste verkenning van de toetsproblematiek bleek het enkel ontwikkelen van een toets die een afspiegeling is van de leercondities onvoldoende. Met een voorafgaande analyse van het leerproces dat plaatsvindt, gelet op de combinatie

van instructie en toetsing en de functies die representaties daarin vervullen, is het onderzoekstraject verbreed met als onderzoeksvraag:

Wat is een effectieve en efficiënte combinatie van instructie en toetsing voor het bereiken van onderwijsdoelstellingen die verbonden zijn met leren door observeren van concreet biologisch materiaal.

Effectiviteit verwijst hierbij naar de mate waarin de onderwijsdoelstellingen worden behaald met als uitgangspunt het te leren construct. Efficiëntie verwijst naar het ontwerpen, ontwikkelen en bewerkstelligen van een instructie-toetscombinatie waarmee de beoogde resultaten worden bereikt tegen zo weinig mogelijke kosten. Beiden zijn onlosmakelijk aan elkaar verbonden en bepalen samen of het eindresultaat al of niet productief is (Molenda, 2009).

Voor het onderzoekstraject is gekozen voor een ontwerpgerichte onderzoeksaanpak om een tweetal redenen. Ten eerste biedt deze de mogelijkheid om binnen de reële onderwijssetting van het dissectiepracticum een educatieve interventie op systematische wijze te ontwerpen, ontwikkelen en evalueren. Ten tweede biedt ontwerpgericht onderzoek de mogelijkheid om de kennis te verruimen over de kenmerken van deze interventie gericht op theorievorming over leren, onderwijzen en toetsen van kennis die wordt ontleend aan realistische objecten. Dit maakt het onderzoek zowel pragmatisch als theoretisch van aard. Allereerst is op pragmatische wijze de toets ontwikkeld waarbij tegelijkertijd de instructie is geoptimaliseerd. Op deze wijze is de instructie-toetscombinatie verbeterd. Daarnaast is getracht om bij het ontwikkelen van de toets algemene richtlijnen te ontleen voor de toetsmethodiek die bijdragen aan de theorievorming over toetsen met visuele representaties. Het actief betrekken van vakdocenten, de faculteitsstaf en gebruikers in het onderzoek heeft het mogelijk gemaakt direct te interveniëren binnen de context waarin het onderzoek zich afspeelt. Hun inbreng heeft bijgedragen aan de verbetering van de instructie-toetscombinatie en het ontwerpen van de nieuwe CBT-toets. Doordat gebruikers binnen de reële context zijn betrokken bij de evaluatie van de toets werd ook informatie gegenereerd over de toepasbaarheid daarvan. Dit gaf informatie over de effectiviteit en efficiëntie van de instructie-toetscombinatie en de ontworpen toets in deze specifieke onderwijssituatie. Voor de onderzoeker bood de ontwerpgerichte onderzoeksaanpak, als bijkomend voordeel, de mogelijkheid om met een diversiteit aan onderzoeksinstrumenten het gehele onderzoekstraject te doorlopen waarbij alle op de interventie gerichte activiteiten tijdens het ontwerpproces empirisch zijn uitgevoerd.

Het gevarieerde onderzoeksinstrumentarium omvatte een literatuuronderzoek, analyse van het instructieobject door dissectie uitgevoerd door de wetenschapper, de toepassing van een curriculum-evaluatiemodel ter analyse van de instructie, intensieve samenwerking met vakdocenten in alle fasen van het onderzoek, semi-gestructureerde interviews met docenten en stafleden van de vakgroep ter evaluatie van het te leren construct, een protocolanalyse bij de studenten ter evaluatie van de instructie-toetscombinatie evenals experimenten met bijbehorende analyse-instrumenten.

7.2 BEANTWOORDING VAN DE ONDERZOEKSVRAGEN

Om de algemene onderzoeksvraag te kunnen beantwoorden is deze uitgewerkt naar specifiekere onderzoeksvragen die zijn onderzocht met twee onderzoekslijnen waarin drie deelonderzoeken zijn uitgevoerd. De eerste onderzoekslijn is gericht op het ontwikkelen van een geschikte instructie-toetscombinatie als alternatief voor de bestaande onderwijssituatie. Het eerste deelonderzoek is de evaluatie en aanpassing van de instructie voor het bereiken van de leerdoelen en de aansluiting op de te ontwikkelen toets. Het tweede deelonderzoek is het ontwerpen en testen van een geschikt toetsconcept. De tweede onderzoekslijn, en derde deelonderzoek, is de vormgeving van de visuele representaties bij de toets in termen van realisme en context.

Instructie-toetscombinatie

De eerste onderzoekslijn leverde een nieuwe instructie-toetscombinatie op waarbij de definitie van het construct is bijgesteld, een CBT toets is ontwikkeld die beantwoordt aan de behoefte van de faculteit Wiskunde en Natuurwetenschappen en de instructie is aangepast aan de nieuwe toetsmethode.

Aanpassing van de instructie

Uit het onderzoek naar de instructie blijkt dat een aangepaste instructie met gewijzigde instructietaken even effectief is als de bestaande instructie. Van oudsher zijn de instructietaken voor dissectiepractica: observatie van het voorsnijden, het zelf uitvoeren van de dissectie en het vastleggen van de geobserveerde structuren in een tekening met hun Latijnse benamingen. De taak om geobserveerde structuren vast te leggen in een tekening is een onderwijs traditie binnen de biologie vanuit de overtuiging dat met deze taak de instructie het effectiefst is (Jones, Reed & Weyers,

1994, 1998; Gaupp, 1904). Het in hoofdstuk 2 gerapporteerde literatuuronderzoek bevestigde dat tekenen inderdaad een effectieve onderwijsmethode is waarbij deze effectiviteit vooral gelegen is in het sturing geven aan het leerproces (Van Meter, Aleksic, Schwartz & Garner, 2006; Van Meter & Garner, 2005; Van Meter, 2001; Lansing, 1984, 1981). Tekenend kan worden beschouwd als een taakgerichte opdracht die door sturing in de selectie van de belangrijke visuele kenmerken de waarnemer dwingt een correcte mentale representatie te construeren tijdens het leerproces. Uit de literatuur blijken ook andere taakgerichte opdrachten effectief zoals het aanbieden van een standaardtekening met de taakgerichte opdracht om de objecten daarin te benoemen. Voor een betere aansluiting van de instructie op de te ontwerpen toetsmethode is voor dit type opdracht gekozen De instructie is aangepast door het tekenen te vervangen door het benoemen van anatomische structuren in een aangeboden standaardtekening en het in groepsverband nabespreken van het geleerde met een poster. In een experiment is onderzocht of de op deze wijze vernieuwde instructie gelijkwaardig is aan de bestaande instructie voor het bereiken van de onderwijsdoelen van het dissectiepracticum anatomie van de rat door deze met elkaar te vergelijken. Het voornaamste resultaat is dat beide instructievormen even effectief bleken. Met dit onderzoeksresultaat is gekozen te investeren in de aangepaste instructiemethode die is afgestemd op de te ontwikkelen toets. De keuze voor de aangepaste instructie levert ook enkele efficiëntie voordelen op. De aangepaste instructie is minder arbeidsintensief doordat de tijdsdruk voor het maken van de tekeningen is komen te vervallen. Studenten hebben meer tijd om actief met elkaar over de dissectielagen te discussiëren en zich de leerstof eigen te maken. Daarnaast waren de studenten ook per instructiedag een uur eerder klaar en deze tijdspanne is benut voor de toetsafname.

Bijstelling van het construct

Een volgend resultaat in het hier gepresenteerde onderzoek is de bijstelling van de constructdefinitie. De constructdefinitie zoals gedefinieerd aan de start van het onderzoek was vooral gebaseerd op kennis over de statische configuratie van de anatomische structuren: het kunnen herkennen van visuele kenmerken naar vorm, kleur en textuur, het kunnen benoemen van visuele kenmerken en het kunnen aangeven hoe de structuren ruimtelijk aan elkaar gerelateerd zijn. Na evaluatie van de constructdefinitie met vakdocenten en onderwijsstaf, blijkt de tot dan toe gehanteerde constructdefinitie niet dekkend te zijn voor wat er daadwerkelijk wordt geleerd gedurende het practicum. De studenten leren ook over de functies van de systemen in een zoogdier en de onderwijsstaf gaf aan dat

ook belangrijk te vinden. De drie kennistypen van Hegarty (2005) blijken toepasselijk met als resultaat een bijgestelde constructdefinitie waarbij de oorspronkelijke definitie is aangevuld. In de bijgestelde constructdefinitie is van belang dat studenten kennis hebben van: de visuele kenmerken van deze structuren (vorm, grootte, kleur, textuur, spatiële verhoudingen), de verbale labels in het Latijn en beschrijvingen van deze structuren evenals de relaties tussen deze structuren op basis van hun functionaliteit en causaal gedrag.

Met deze aanpassingen in de instructie en bijstelling van de constructdefinitie is de toetsontwikkeling voortgezet.

Constructie van de CBT-toets

Een derde resultaat van het onderzoek is een CBT-toets die een adequate weerspiegeling blijkt van het te leren construct. Daartoe is eerst in de literatuur gezocht naar toetsvormen die de resultaten van het visueel leerproces, het geleerde construct, kunnen toetsen op het behalen van de gestelde leerdoelstellingen. Uit de literatuur in hoofdstuk 2 blijkt weinig informatie voorradig over visueel toetsen en de effectiviteit daarvan in het toetsen van het geleerde construct. Bij benadering is het PSE-toetsprogramma van Dwyer (1976-2007) het meest uitgewerkt met visuele representaties hoewel deze is beperkt tot het toetsen van de effectiviteit van verbale instructies, tekstueel of mondeling, die worden ondersteund door representaties. Voor wat betreft de toetsconstructie blijkt uit de in hoofdstuk 2 gepresenteerde literatuur dat een visuele toets een visueel leerproces beter beoordeelt en de retentie verstevigt dan een toets met een andere modaliteit zoals een tekstuele toets (Braden, 1996; Dwyer, 1978, Dwyer & De Melo, 1983; Moore & Dwyer, 1994). Voor de ontwikkeling van de toets in het hier gepresenteerde onderzoek is gekozen voor de ontwikkeling van een CBT-toets met als voorname redenen de mogelijkheid tot een eenduidige, geautomatiseerde scoringsmethodiek, het aanbieden van een diversiteit aan visuele representatievormen die het leerobject in de leersituatie benaderen en de mogelijkheden in het aanbieden van toetstaken. Deze mogelijkheden zijn in andere toetsvormen beperkt, ook blijkend uit de papierentoets die in het hier gepresenteerde onderzoek is ontwikkeld als een eerste toetsversie ten behoeve van de vergelijking van de instructies in hoofdstuk 3. In samenspraak met de vakdocenten is deze eerste papieren toetsversie met een stappenplan (Van Oosterveld & Vorst, 1998) doorontwikkeld naar een CBT-toets en in de loop van het onderzoek bijgesteld voor de verbetering van de kwaliteit. In tegenstelling tot

de papieren toetsversie is de CBT- toets niet beperkt tot enkel items met een herkenningstaak maar uitgebreid met items met een benoemingstaak en items met een relatietask waarin constructieopdrachten zijn verwerkt. Deze toetstaken weerspiegelen samen het gedefinieerde construct.

Involed van factoren die de visualisatie(vorm) beïnvloeden: realisme en context

In de tweede onderzoekslijn is nagegaan welke representatievorm het construct bij benadering het beste toetst. In de literatuur in hoofdstuk 2 wordt het realiteitsgehalte van een representatie als belangrijke visueel kenmerk meerdere malen benoemd. Naar aanleiding van de literatuur is in hoofdstuk 2 een rangschikking van varianten in visualiteitsvormen gepresenteerd waarin expliciet de relatie wordt gelegd tussen de functie en de bijdrage van een visualiteitsvorm en typen te verkrijgen kennis (tabel 2.1). Daarop voortredenerend blijkt de inbreng van de variabelen realisme en context van belang. Realisme is de variatie in de hoeveelheid details die in elk (deel)object wordt weergegeven variërend van abstract schematisch tot een gedetailleerde fotoversie. De context is variatie in het realiteitsgehalte door de representatie van de (deel)objecten afzonderlijk van elkaar of binnen het geheel van onderlinge relaties en samenhang.

Uit de literatuur blijken er geen eenduidige resultaten te zijn over de hoeveelheid realiteit (realistisch versus schematisch) die noodzakelijk is voor gebruik in visuele representaties. Algemene conclusie is dat realisme een relevant aspect is in het leerproces maar met verschillende effecten blijkend uit het gebrek aan eenduidige onderzoeksresultaten in de literatuur. Evenmin biedt de literatuur duidelijkheid over de relatie tussen realisme en het toetsen van een geleerd construct. Voor het hier gepresenteerde onderzoek is van belang na te gaan welke visualiteitsvorm, varianten op realisme, het meest bijdraagt aan het toetsen van het construct. Realistische foto's benaderen het meest het reële object maar zijn kostbaar in de ontwikkeling van de CBT-toets. Contourschema's daarentegen zijn versimpelde weergaven van het reële object maar kosten minder in de ontwikkeling van de CBT-toets. Om die reden is besloten realisme te onderzoeken vanuit een gedetailleerde weergave, een foto, versus een gesimplificeerde weergave, een schematische tekening met alleen de contouren van de af te beelden objecten, met als specifieke onderzoeksvraag:

Meet een toets met foto's hetzelfde visuele construct als een toets met schema's?

De context is vooral gericht op de wijze waarop objecten zich op een bepaalde wijze tot elkaar verhouden door onder meer vorm, positie, grootte, textuur en andere kenmerken die gezamenlijk een Gestalt vormen. Volgens de Gestaltprincipes (Arnheim, 1974) is perceptie gericht op het begrijpen van definieerbare vormen en patronen waarbij de Gestalt wordt gevormd door de groepering van stimuli tot één geheel. Kenmerken van een Gestalt zijn: (a) dat het geheel eerder wordt waargenomen dan de delen waaruit de Gestalt is opgebouwd, (b) dat als één element binnen de Gestalt verandert, deze meteen als totaliteit verandert en (c) dat de Gestalt altijd meer is dan de som van de aparte elementen waaruit deze is samengesteld. In de literatuur is, in de lijn van de hier gepresenteerde onderzoekssituatie, geen vergelijkbaar onderzoek gevonden naar de effectiviteit van visuele representaties als de Gestalt in een representatie onvolledig is of het referentieobject niet geheel weergeeft. Daarmee is er ook geen uitsluitsel over de noodzaak tot aan- of afwezigheid van context in relatie tot het te toetsen construct.

Dit heeft geleid tot het besluit ook context te onderzoeken, met als onderzoeksvraag:

Meet een toets met context hetzelfde visuele construct als een toets zonder context?

Voor het beantwoorden van deze specifieke onderzoeksvragen is de ontwikkelde CBT-toets vertaald naar vier toetsversies met elk een eigen representatievorm: foto met context, foto zonder context, schema met context en schema zonder context.

Uit de onderzoeksresultaten, gepresenteerd in hoofdstuk 6, blijkt dat bij vergelijking van de vier toetsversies deze toetsen vergelijkbaar zijn waarbij het merendeel van de items in de toetsvormen hetzelfde functioneert. Voor het hier gepresenteerde onderzoek is de conclusie dat de onderzoeksvragen positief kunnen worden bevestigd: de vier toetsvormen meten hetzelfde visuele construct in de verschillende representatievormen. Visualiteit in termen van realisme, foto of schema, en context, al of niet, heeft geen significante invloed op het toetsen van het construct. Dit biedt voor toekomstig gebruik de mogelijkheid voor de vakdocenten om de items van de vier toetsvormen samen te voegen tot één dataverzameling en een gecombineerde toets samen te stellen van de vier visualiteitsvormen waarmee hetzelfde construct wordt gemeten. Hierbij dient te worden opgemerkt dat de toetsvormen met foto's, met of zonder context, blijkens de IRT-analyse in hoge mate met elkaar overeenkomen (tabel 6.6).

Dit onderzoeksresultaat betekent niet dat visualiteit helemaal niet van invloed is bij het meten van het construct. Een aantal items uit de toetsversies blijkt namelijk verschillend te functioneren over de toetsen heen, waarbij uit de onderzoeksresultaten blijkt dat er wel degelijk sprake is van een significant verschil voor de variabele realisme tussen foto en schema (zie tabel 6.9). Deze items lenen zich in principe voor nadere studie om te bezien of op grond van de relatie tussen de visualisatie, de inhoud van de plaatjes en de toetstaak zou kunnen worden begrepen waarom deze items verschillend functioneren in de verschillende toetsversies. Zoals al in hoofdstuk 6 opgemerkt zouden daarvoor meer data beschikbaar moeten zijn uit meerdere toetsafnames om eerst te weten of de gevonden effecten structureel zijn.

Moeilijkheidsgraad van de toetstaken

Een vijfde resultaat van het hier gepresenteerde onderzoek is de gevonden moeilijkheidsgraad tussen de toetstaken, een bijproduct van het onderzoek. Uit de onderzoeksresultaten blijkt dat items met herkenningstaken gemakkelijker zijn dan items met benoemings- en relatietaken. Ook zijn items met benoemingstaken gemakkelijker dan items met relatietaken. Dat de items met herkenningstaken gemakkelijk zijn, is een in de literatuur bekend feit, zie bijvoorbeeld de in hoofdstuk 2 genoemde bevinding van Levie (1987) dat lerenden in veel voorkomende gevallen hoog scoren op herkenningstaken. Herkennen is ook de zwakste toetstaak in elke representatievorm. De aanwezigheid van andere toetstaken is daarom noodzakelijk om de toets onderscheidend te doen zijn in het meten van het te toetsen construct. Door in de ontwikkeling van de CBT-toets met verschillende itemvormen te werken (meerkeuze-items en de CFR-items), is een brede toets geconstrueerd met meerdere toetstaken waardoor het meten van het construct niet beperkt is tot, en afhankelijk is van, het kunnen herkennen. Dat bleek voor het hier gepresenteerde onderzoek van meerwaarde. Aangezien dit resultaat een bijproduct is van het onderzoek wordt aanbevolen in vervolgonderzoek hier specifiek aandacht aan te besteden.

7.3 EINDRESULTAAT VOOR DE FACULTEIT WISKUNDE EN NATUURWETENSCHAPPEN

Met de ontwikkeling van een CBT-toets die objectief en eenduidig wordt gescoord, die aansluit op het leerproces en die studenten in een kort tijdsbestek voorziet van feedback wordt beantwoord aan de opdracht van de faculteit. De

definitie van het te leren construct is geëvalueerd en verbeterd. De verandering houdt in dat naast het herkennen en benoemen van anatomische structuren, het leggen van relaties tussen anatomische objecten op basis van hun functie of hun rol in een fysiologisch systeem, onderdeel is geworden van het te toetsen construct. De toetsvorm is daarop aangepast en daarmee verbeterd in vergelijking tot de vorige toetsvorm waarin tekeningen werden nagekeken op alleen de factoren herkennen en benoemen.

De in het onderzoek ontwikkelde instructie sluit aan op de toets en is in een korter tijdsbestek te doorlopen. Deze tijdsbesparing is in termen van efficiëntie een goed nevenresultaat daar de tijdswinst in de instructie kan worden ingezet voor de toetsafname. Bovendien is bewerkstelligd dat de tijdsdruk in de instructie is verminderd.

Voor toekomstig gebruik van het toetsprogramma is de toets geschikt. Als voor daadwerkelijk gebruik van de instructie-toetscombinatie wordt besloten, dan is het aan te bevelen om gebruik te maken van de suggesties die in hoofdstuk 6 zijn besproken voor het werken met een samengestelde toets met een mix van de verschillende representatievormen. Ook wordt aanbevolen om de implementatie zo vorm te geven dat de evaluatie van de toets structureel in de toetsprocedure wordt ingebouwd om het databestand over de items verder op te bouwen om hun eigenschappen verder te onderzoeken.

7.4 SLOTOPMERKING

Aan de primaire opdracht van de faculteit is voldaan. De opbrengst in de twee onderzoekslijnen, is niettemin beperkt. Voor deze ene situatie is een bruikbare instructietoetscombinatie ontwikkeld. Over de toepasbaarheid buiten de context van dit onderzoek bestaat geen duidelijkheid. Hopelijk wordt op dit punt in vergelijkbare projecten meer ervaring opgebouwd, waarbij de resultaten en de werkwijze in dit onderzoek inspiratie kunnen bieden. Op het gebied van de werking van verschillende typen visualisatie, is de bijdrage beperkt tot de constatering dat met de in het onderzoek gebruikte varianten geen verschillen ontstaan in de te meten constructen. Wel is er sprake van verschillende niveaus van moeilijkheid maar de mate waarin dit structureel is, behoeft verder onderzoek.

Als de faculteit besluit om met de ontwikkelde instructie-toetscombinatie verder te gaan, dan ontstaat wel een testomgeving waarin ook op dit punt interessant vervolgonderzoek mogelijk wordt.

Tenslotte: Sinds de start van dit project, zijn op het gebied van toetsen en van visualisaties en daarvoor beschikbare technologieën veel ontwikkelingen geweest en voortdurend gaande. Het kan uitdagend zijn om in de bedoelde testomgeving ook ruimte te bieden om te onderzoeken wat de nieuwe mogelijkheden daar te bieden hebben.

REFERENTIES

- Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, 7, 255-278.
- Ackerman, T. A. (1996). Graphical representation of multidimensional item response theory analyses. *Applied Psychological Measurement*, 20, 311-330.
- Ainsworth, S. E. (2006). DeFT: A conceptual framework for learning with multiple representations. *Learning and Instruction*, 16(3), 183-198.
- Akker, J. van den, McKenney, S., Nieveen, N., & Gravemeijer, K. (2006). Introducing educational design research. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 3-7). London: Routledge.
- Alesandrini, K. L. (1984) Pictures and adult learning. *Instructional Science*, 13, 63-77.
- Alesandrini, K. (1981). Pictorial-verbal and analytic-holistic learning strategies in science learning. *Journal of Educational Psychology*, 73(3), 358-368.
- Anglin, G. J., Vaez, H., & Cunningham, K. L. (2004). Visual representations and learning: The role of static and animated graphics. In D. Jonassen, (Ed.), *Handbook of research on educational communications and technology* (2nd ed., pp. 865-916). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Arnheim, R.(1986). *New essays on the psychology of art*. Los Angeles, CA: University of California Press
- Arnheim, R. (1974). *Art & visual perception: A psychology of the creative eye*. Berkeley: University of California Press.
- Baddeley, A.D. (1992). Working memory. *Science*. 255, pg 556-559
- Bejar, I. I. (1991). A methodology for scoring open-ended architectural design problems. *Journal of applied psychology*, 76, 522-532.
- Bejar, I. I. & Braun, H. I. (1999) *Architectural simulations: From research to implementation final report to the national council of architectural registration boards*. Princeton, New Jersey: Educational testing service.
- Behrmann, M. and Moscovitch, M. (2001). Face recognition. In F. Boller and J. Grafman (Eds.) *Handbook of Neuropsychology*. 2nd edition. Elsevier Science, North Holland, 181-206.
- Bennett, R. E., & Bejar, I. I. (1998). *Validity and auto- mated scoring: It's not only the scoring*. Educational Measurement: Issues and Practice, 17 (4), 9-16.
- Bloom, B. S. (1994). Reflections on the development and use of the taxonomy. In Rehaeg, Kenneth J.; Anderson, Lorin W.; Sosniak, Lauren A. "Bloom's taxonomy: A forty-year retrospective". Yearbook of the National Society for the Study of Education (Chicago: National Society for the Study of Education) 93 (2).

- Borges, M. A., Stepnowsky, M. A., & Holt, L. H. (1977). Recall and recognition of words and pictures by adults and children. *Bulletin of the Psychonomic Society*, 9(2), 113-114.
- Braden, R. A. (2004). Visual literacy. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (2nd ed). Lawrence Erlbaum Associates.
- Braden, R.A. (1996). Visual literacy. In D.H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 491-520). New York: Simon & Schuster Macmillan
- Brink, W. P. van den, & Mellenbergh, G. J. (2003). *Testleer en testconstructie*. Amsterdam: Boom.
- Britton, L. A., & Wandersee, J. H. (1997). Cutting up text to make moveable, magnetic diagrams: A way of teaching and assessing biological processes. *The American Biology Teacher*. 59: 288-291.
- Brody, P. J. (1982). Affecting instructional textbooks through pictures. In D. H. Jonassen (ed.), *The Technology of Text*, 301-316.
- Bruce, Green, & Georgeson (1997). *Visual perception: physiology, psychology and ecology*. London: Psychology Press.
- Butler, L., & McKelvie, S. J. (1985). Processing of form: further evidence for the necessity of attention. *Perceptual and Motor Skills*, 61, 215-221.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Campbell, D. T. & Stanley, J. C. (1981). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin
- Cave, C. B., & Kosslyn, S. M. (1993). The role of parts and spatial relations in object identification. *Perception*, 22, 229-248.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4): 293-332.
- COTAN (2010). *Beoordelingssysteem voor Kwaliteit van Toetsen*. Uitgave van het Nederlands Instituut voor Psychologen (NIP).
- Cronbach, L. J. & Meehl, P. E. (1955) Construct validity in Psychological tests. *Psychological Bulletin*, 52, 281-302, p4 & p27
- Dean, R. S., & Kulhavy, R. W. (1981). Influence of spatial organization on prose learning. *Journal of Educational Psychology*, 73, 57-64
- Dempsey, B. C. & Betz, B. J. (2001) Biological Drawing: A Scientific Tool for Learning. *The American Biology Teacher*, 63 (4) (2001), pp. 271-279
- Downing, S. M. (2006) Selected response item formats in test development. Downing, S. M., and T. M. Haladyna, eds. (2006). *Handbook of Test Development*. Mahwah, N.J.: Lawrence Erlbaum Associates
- Downing, S. M., and T. M. Haladyna, eds. (2006). *Handbook of Test Development*. Mahwah, N.J.: Lawrence Erlbaum Associates
- Drenth, P. J. D., & Sijtsma, K. (1990). *Testtheorie: Inleiding in de theorie van de psychologische test en zijn toepassingen*. Houten: Bohn Stafleu Van Loghum.
- Duchastel, P., & Waller, R. (1979). Pictorial illustration in instructional texts. *Educational technology*, 19(11), 20-25
- Dwyer, F. M. (1994). One dimension of visual research: a paradigm and its implementation. In Moore, D. M., & Dwyer, F. M. (Ed.), *Visual literacy: A spectrum of visual learning*. New Jersey: Educational Technology Publications.

- Dwyer, F. M (1978). *Strategies for improving visual learning*. State College, Pa: Learning Services
- Dwyer, F. M., & De Melo, H. (1983) The effect of visual testing in assessing the instructional potential of variables associated with visualized instruction. *Journal of Instructional Psychology*, 10, 126-140
- Gagné, R.M. (1985). *The Conditions of Learning and Theory of Instruction (4th Edition)*. New York: CBS College Publishing.
- Gagné, R.M. (1977). *The conditions of learning*, 3rd edition. New York: Holt, Rinehart and Winston
- Gagne, R. M. (1972) Domains of Learning. *Interchange*, 3, 1-8
- Gagné, R. M. (1965). *The conditions of learning*. New York: Holt, Rinehart, & Winston
- Gaupp, E., (1904). *Anatomic des Frosches*. Braunschweig, Vieweg & Sohn.
- Geerts, G. & Boon, C.A. den (1999). *Van Dale: Groot Woordenboek der Nederlandse Taal: 3 Dl.* (13e dr.). Utrecht: Van Dale Lexicografie
- Gibson, J. J. (1972). A theory of direct visual perception. In J. Royce, & W. Rozenboom (Eds.), *The psychology of knowing*. New York: Gordon & Breach.
- Glas, C. A. W. (2010). Item parameter estimation and item fit analysis. In W. J. van der Linden and C. A. W. Glas (Eds.), *Elements of Adaptive Testing* (pp. 269-288). New York, NJ: Springer.
- Glas, C. A. W. (1999). Modification indices for the 2-pl and the nominal response model. *Psychometrika*, 64, 273-294
- Goldstein, A.G. (1975). Recognition of inverted photographs of faces by children and adults. *Journal of Genetic Psychology*, 127, 109-123.
- Gombrich, E. H. (1969). *Art and illusion*. Princeton: Princeton university press
- Graaff De, J. & Verhagen, P.W., (1997, Feb). An interactive correction system for visual testing of anatomical drawings. Paper presented at the 1997 National Convention of the Association for Educational Communications and Technology (AECT), Albuquerque, NM, February 12-16, 1997
- Groot, A. D. de (1961). *Methodologie*. Den Haag: Mouton.
- Hegarty M. (2005.) Multimedia learning about physical systems. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 447-465). Cambridge: Cambridge University Press.
- Hegarty, M. & Steinhoff, K. (1997) Individual differences in use of diagrams as external memory in mechanical reasoning. *Learning and individual differences*. 9, 19-42
- Heiser, J. and Tversky, B. (2002). Diagrams and descriptions in acquiring complex systems. *Proceedings of the Cognitive Science Society*. Erlbaum: Hillsdale, N.J
- Hortin, J. A. (1994). Theoretical foundations of visual learning. In D. M. Moore, & F. M. Dwyer (Ed.), *Visual literacy: A spectrum of visual learning*. New Jersey: Educational Technology Publications.
- Jonassen, D. H., Tessmer, M. & Hannum, W. H. (1999). *Task analysis methods for instruction*. Mahwah, New Jersey: Lawrence Erlbaum Associates
- Jones, A., Reed, R., & Weyers, J. (1994, 1998). *Practical skills in biology*. New York: Longman Scientific & Technical.
- Kinzie, M.B., Strauss, R., & Foss, J. (1993). The effects of an Interactive Dissection Simulation on the Performance and Achievement of High School Biology Students. *Journal of Research in Science Teaching* 30(8).

- Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- Knowlton, J. (1966). On the definition of 'picture'. *AV Communication review*, 14, 157-183.
- Lansing, K.M. (1984) The effect of drawing on the development of mental representations: a continuing study. *Studies in art education*, 25 (3), 167-175
- Lansing, K. M. (1981). The effect of drawing in the development of mental representations. *Studies in Art Education*, 22(3), 15-23.
- Levie (1978) quoted in Anglin, Vaez & Cunningham pg 867.
- Levie, W. H. (1987). Research on pictures: A guide through the literature. In D. M. Willows, & H. H. Houghton (Eds.), *The psychology of illustration: Basic research* (Vol. 1). New York: Springer-Verlag.
- Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. *EJTC*, 30, 195-232.
- Loftus, G. R., & Bell, S. M. (1975). Two types of information in picture memory. *Journal of Experimental Psychology: Human Learning and Memory*, 104, 103-113.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.
- Martinez, M. E. (1998). *A taxonomy of methods for demonstration of proficiency in a figural medium*. College Park, MD: Clearinghouse on Assessment and Evaluation. (ERIC Document Reproduction Service No. ED 423 241).
- Martinez, M. E., & Katz, I. R. (1992). *Cognitive processing requirements of constructed figural response and multiple-choice items in architecture assessment*. Princeton, New Jersey: Educational testing service.
- Martinez, M.E. (1990). *A comparison of multiple-choice and constructed figural response items*. Princeton, New Jersey: Educational testing service.
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*. Cambridge, New York: Cambridge University Press.
- Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review*, 8, 357-371.
- Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90, 312-320.
- McKenney, S., Nieveen, N., & Akker, J. van den, (2006). Design research from a curriculum perspective. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 67-90). London: Routledge.
- Merrill, M.D. (1983). Component Display Theory. In C. Reigeluth (ed.), *Instructional Design Theories and Models*. Hillsdale, NJ: Erlbaum Associates.
- Messick, S. (1993). *Foundations of validity: Meaning and consequences in psychological assessment*. Educational Testing Service. Princeton, New Jersey
- MIRT (2010). *Multidimensional Item response Theory*. Public Domain Software. <http://www.utwente.nl/gw/omd/Medewerkers/medewerkers/glas/>
- Molenda, M. (2009) Instructional technology must contribute to productivity. *Journal of computing in higher education*. Vol. 21 n1, pg 80-94. April 2009
- Moore, D.M. & Dwyer, F.M. (Ed.) (1994) *Visual literacy: a spectrum of visual learning*. New Jersey: Educational Technology Publications

- Mory, E. H. (1996). Feedback research. In D. H. Jonassen (ed.), *Handbook of research for educational communications and technology* (pp. 919-956). New York: Simon & Schuster Maxmillan.
- Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (ed.), *Handbook of research for educational communications and technology* (2nd edition, pp. 745-783). Mahwah, NJ: Lawrence Erlbaum Associates.
- Moscovitch, M., & Moscovitch, D. A. (2000). Super face-inversion effects for isolated internal or external features, and for fractured faces. *Cognitive Neuropsychology*, 17(1-3), 201-219.
- Nelson, T. O., Metzler, J., & Reed, D. A. (1974). Role of detail in the long-term recognition of pictures and verbal descriptions. *Journal of Experimental Psychology*, 102, 184-186.
- Oosterveld, P. & Vorst, H.C.M. (1998). Methoden van vragenlijstconstructie [Methods of questionnaire construction]. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden*, 51, 11-27.
- O'Neill, K., & Folk, V. (1996, April). Innovative CBT item formats in a teacher licensing program. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.
- Paivio, A (1986). *Mental representations: a dual coding approach*. Oxford. England: Oxford University Press
- Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory and Cognition*, 3(5), 519-526.
- Parshall, C., Davey, T., & Pashley, P. (2000). Innovative itemtypes for computerized testing. In W. J. Linden, van der, & C. A. W. Glas, (Ed.), *Computerized adaptive testing: Theory and practice*. Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Parshall, C. G., Harmes, J. C., Davey, T., & Pashley, P. J. (2010). Innovative item types for computerized testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 215-230). New York, NY: Springer.
- Pettersson, R. (1993) *Visuals for information, research and practice*, 2de ed. Englewood cliffs, NJ: Educational Technology
- Plomp, Tj. (2009). Educational design research. In Tj. Plomp, & N. Nieveen. *An introduction to educational design research*. Enschede: SLO- Netherlands institute for curriculum development.
- Randhawa, B. S. (1978). *Visual learning, thinking and communication*. New York, London: Academic Press / 1978.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Reigeluth, C. M. (1994). The imperative for systemic change. In C. M. Reigeluth & R. J. Garfinkle (Eds.), *Systemic change in education* (pp. 3-11). Englewood Cliffs, NJ: Educational Technology Publications.
- Ritchey, G. H. (1982). Pictorial details and recall in adults and children. *Journal of Experimental Psychology*, 8(2), 139-141.
- Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 49-69). Cambridge: Cambridge University Press.

- Schnotz, W. (2002). Towards an integrated view of learning from text and visual displays. *Educational Psychology Review*, 14(2), 101-120
- Seels, B. A. (1994). Visual literacy: The definition problem. In D. M. Moore, & F. M. Dwyer (Ed.) *Visual literacy: A spectrum of visual learning*. New Jersey: Educational Technology Publications
- Serici, S. G. & Zenisky, A. (2006) Innovative item formats in computer-based testing: in pursuit of improved construct representation. In Downing, S.M. & Haladyna, T.M.. (Eds.). *Handbook of Test Development* (pp. 329-347). Mahwah, NJ: Lawrence Erlbaum Associates.
- Smith, P.L. & Ragan, T.J. (1993). *Instructional Design*, New York: Macmillan (1999 edition by Wiley & Sons)
- Smith M. K., Wood W. B., Adams W. K., Wieman C., Knight J. K., Guild N., Su T. Tt (2009). Why peer discussion improves student performance on in-class concept questions. *Science* 323, 122-124.
- Stake, R. E. (1967). The Countenance of Educational Evaluation. *Teachers College Record*, 68, 523-540
- Stein, M., & Power, B. (1996). Putting art on the scientist's palette. In Hubbard, R. S., & Ernst, K. (eds.), *New Entries: Learning by Writing and Drawing*, Heinemann, Portsmouth, NH.
- Stern, R. C., & Robinson, R. S. (1994). Perception and its role in communication and learning. In D. M. Moore, & F. M. Dwyer (Ed.), *Visual literacy: A spectrum of visual learning*. New Jersey: Educational Technology Publications.
- Sijtsma, K., & Molenaar I. W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks: Sage.
- Swanborn, P. G. (1993). *Methoden van social-wetenschappelijk onderzoek*. Amsterdam: Boom Meppel
- Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19-30). Cambridge: Cambridge University Press.
- Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990) Cognitive load and selective attention as factors in the structuring of technical material. *Journal of Experimental Psychology General*, 119, 176-192
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The quarterly journal of experimental psychology*, 46A(2), 225-245.
- Van der Linden, W. J., & Hambleton, R. K. (1997) (Eds.). *Handbook of modern item response theory*. New York, NJ: Springer.
- Van Meter, P., Aleksic, M., Schwartz, A., & Garner (2006). Learner-Generated Drawing as a Strategy for Learning from Content Area Text. *Contemporary Educational Psychology*, 31, 142 – 166.
- Van Meter, P. & Garner, J. (2005). The promise and practice of learner-generated drawing: literature review and synthesis. *Educational Psychology Review*, 17, 285 – 325.
- Van Meter, P. (2001). Drawing construction as a strategy for learning from text. *Journal of Educational Psychology*, 69, pp. 129 – 140.
- Waters, J.R., P. Van Meter, W. Perrotti, S. Drogo, and R. Cyr. (2005). Cat dissection vs. sculpting human structures in clay: an analysis of two approaches to undergraduate human anatomy laboratory education. *Advances in Physiology Education* 29(1):27-34

- Wang, M. C., G. D. Haertel, and H. J. Walberg. (1990). "What Influences Learning? A Content Analysis of Review Literature." *Journal of Educational Research* 84: 30-43.
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63(3), 249-294.
- Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, 24, 345-376.

NEDERLANDSE SAMENVATTING

Dissectie en observatie: Visueel toetsen en visueel leren

Visueel leren, in brede zin, is kennis dat wordt opgedaan in een leeromgeving met visuele representaties. Onderwijssituaties waarin de lerende zich de eigenschappen van een concreet studieobject eigen maakt door observatie zijn daar voorbeelden van, zo ook dissectiepractica zoals het practicum 'Anatomie rat'. In deze onderwijssituatie wordt een zoogdier laag voor laag voorzichtig ontleed waarbij observatie en identificatie van de verschillende anatomische structuren de voornaamste handelingen zijn ten behoeve van het leerproces. Het toetsen van kennis opgedaan met visueel leren blijkt het best met visuele representaties, visueel toetsen (Braden, 1996; Dwyer & Moore, 1994). Het is zinvoller een visuele instructie te toetsen in dezelfde modaliteit dan een andere. Dit is voor het practicum 'Anatomie rat' niet mogelijk en wordt opgelost door tekeningen, die studenten op basis van specifieke aanwijzingen maken tijdens het leerproces, te beoordelen. Deze visuele toetsmethodiek is om verschillende redenen niet wenselijk zoals de aanname dat de gemaakte tekening het bewijs is van de kennis die de persoon beheerst over het studieobject. De objectiviteit van de beoordeling is een zorgpunt daar deze handmatig gebeurt met meerdere beoordelaars evenals het lange tijdsbestek tussen instructie en beoordeling waardoor het effect van feedback gering is.

Het onderzoek is gestart om deze problemen het hoofd te bieden door de ontwikkeling van een visuele CBT toets die aansluit op het leerproces, objectief is en in een relatief kort tijdsbestek feedback geeft. Bij nadere bestudering van de problematiek bleek enkel de ontwikkeling van een toets onvoldoende door gebrek aan kennis over de precieze onderwijsdoelen -het aan te leren construct- in relatie tot visueel leren en visueel toetsen. Het onderzoekstraject is verbreed door de onderwijssituatie, vanuit een ontwerpgerichte onderzoeksaanpak, als geheel te bestuderen uitgaand van een samenspel tussen instructie en toetsing met als onderzoeksvraag: Wat is een effectieve en efficiënte combinatie van instructie en toetsing voor het bereiken van onderwijsdoelstellingen die verbonden zijn met leren door observeren van concreet biologisch materiaal.

Er is een vooronderzoek uitgevoerd met daarin literatuur onderzoek dat resulteerde in een conceptueel kader. Vervolgens zijn twee afzonderlijke onderzoekslijnen uitgevoerd die gezamenlijk bijdragen aan het beantwoorden van de onderzoeksvraag. De voornaamste onderzoekslijn is de ontwikkeling van een instructie-toets combinatie die dekkend is voor het construct. de tweede onderzoekslijn is gericht op de meest geschikte representatievorm voor het toetsen van het construct op basis van twee factoren: mate van realisme in de representatie en het al of niet aanwezig zijn van contextuele informatie.

In een eerste experiment is gezocht naar een voor deze onderwijssituaties geschikte instructie-toets combinatie. Daarbij is de bestaande instructie vergeleken met een aangepaste instructie waarin niet meer werd getekend en op andere wijze moest worden beoordeeld. Dit betekende ook de ontkoppeling tussen instructie en toetsing daar beoordelen van tijdens de instructie gemaakte tekeningen, als toetsinstrument, niet meer mogelijk was. Voor de toetsing is een papieren toetsversie ontwikkeld die elektronisch werd verwerkt waardoor de beoordeling eenduidig werd. De instructie is verbeterd door de taakgerichtheid van de practicumtaken te wijzigen. Het maken van tekeningen is vervangen door het aanbieden van tekeningen waaraan de anatomische namen moesten worden toegevoegd. De studenten ontvingen tijdens het practicum feedback door het in de groep bespreken van de eigen ingevulde tekeningen. Bij experimentele vergelijking van de beide instructievormen blijken deze even effectief en is voor het verdere onderzoek gekozen voor de aangepaste instructievorm. Deze is vanwege de aanpassingen efficiënter door tijdswinst in de instructie -omdat er niet meer wordt getekend- en de inkorting van de beoordelingstijd -omdat de tekeningen niet worden beoordeeld-, en door snellere feedback naar de studenten doordat de toets in korte tijd wordt verbeterd en teruggekoppeld.

De papierentoets is vervolgens geëvalueerd en deze bleek niet dekkend voor het geleerde construct. Dit leidde tot een evaluatie en herdefiniëring van het construct verdeeld in drie deelconstructen. In de operationalisatie naar een CBT toets leidde dat tot drie toetstaken: het kunnen herkennen van anatomische objecten aan de visuele kenmerken, het kunnen benoemen van deze objecten in het Latijns en het kunnen leggen van relaties tussen de structuren op basis van hun functionaliteit en causaal gedrag. Het resultaat is een geautomatiseerde toets met 55 toetsitems. Om de onderzoekslijn naar de meest geschikte representatievorm voor het toetsen van het construct te kunnen onderzoeken zijn de factoren realisme en context

geoperationaliseerd met als resultaat vier toetsversies waarvan (1) toets Foto met context, (2) toets Foto zonder context, (3) toets Schema met context en (4) toets Schema zonder context.

In een experiment is deze eerste CBT toets geëvalueerd ter beoordeling of de ontwikkelde instructie-toets combinatie dekkend is voor het construct. Ook is verkend of in deze vorm uitspraken mogelijk zijn over de meest geschikte representatievorm met de vier toetsversies. Uit de resultaten van het experiment blijkt dat de toets op zowel algemeen als op de taakniveau kan worden verbeterd. De betrouwbaarheid van enkele toetsonderdelen in de vier toetsversies, vooral herkennen, blijkt onvoldoende en diverse items bleken zwak en niet onderscheidend. In totaal dient 33% van de items te worden gewijzigd. Ook een eerste variantieanalyse op de toetsversies om mogelijke (hoofd)-effecten op te sporen voor de factoren realisme en context leverde geen effecten op.

De evaluatie van de CBT toets heeft geresulteerd in een bijgestelde toets waarmee het experiment nogmaals is uitgevoerd met vier toetsversies. Uit de resultaten blijkt een duidelijke verbetering van de toetskwaliteit in vergelijking met de vorige toets, met als zwak onderdeel de toetstaak herkennen voor de toetsversies met schema's. Bij vergelijking van de vier toetsversies blijken deze globaal dezelfde constructen te meten hoewel een aantal items verschillend functioneren over de toetsvormen. De toetsversies blijken goed te equivaleren. Voor wat betreft de representatievorm is zijn er hoofdeffecten gevonden van de factor realisme voor de toets als geheel en voor de toetstaken in het herkennen van objecten en in het leggen van relaties tussen objecten. In alle gevallen is het gevonden hoofdeffect in het voordeel van schema's.

Het onderzoek als geheel heeft geleid tot het opleveren van een instructie-toetscombinatie dat geschikt is voor gebruik in het onderwijs en beantwoord aan de problematiek die de aanleiding was voor het onderzoek. Er is een CBT toets ontwikkeld dat dekkend is voor het geleerde construct, objectief is door een eenduidige automatische beoordeling en direct voorziet in feedback. De meeropbrengst is de kwaliteitsverbetering van de instructie die in de aangepaste versie minder tijdrovend is en waarbij gericht aandacht wordt besteed aan feedbacktijdens de instructie. Voorts sluiten instructie en toets goed op elkaar aan met een duidelijke scheiding tussen de instructie en de toetsing hetgeen in de tekeninstructie niet het geval is. Voor wat betreft de representatievorm is er meer onderzoek nodig om daar uitsluitsel over te geven.

Het onderzoek zoals nu opgezet bevatte weinig kandidaten die deelnamen. Door de toets meerdere keren af te nemen, ontstaat er meer volume aan data die bij analyse meer inzichtelijk maakt welke representatievormen geschikt zijn voor welk type toetstaak.

Voor wat betreft de implementatie van de toets zijn er items gedefinieerd die voor verbetering vatbaar zijn. Echter dat verschilt per toetsversie. Het is aan de docenten de items die niet onderscheidend zijn te verbeteren en om een geschikte toets samen te stellen. Door dat de toetsen equivalent zijn is het mogelijk om met de geschikt bevonden-items een dataverzameling te creëren waarmee een toets kan worden samengesteld.

ENGLISH SUMMARY

Dissection and Observation: Visual testing and visual learning

Visual learning, in the broadest sense, is about acquiring knowledge in a learning environment with visual representations. Examples of this are educational situations in which the learner acquires knowledge of the characteristics of a concrete study object through observation, such as dissection practicums. Case in point: the practicum “Rat anatomy”. During this practicum, a mammal is carefully dissected layer by layer, whereby observation and identification of the various anatomical structures are the primary means of learning. The assessment of knowledge acquired through visual learning appears to be done best with visual representations, the so-called *visual testing* (Braden, 1996; Dwyer & Moore, 1994). It is thought to be more sensible to use the same modality for the visual testing as was used for the visual instruction, but this is not feasible for the practicum “Rat anatomy”. Instead, for this educational situation, the students’ knowledge of the subject is assessed based on the drawings made by them during the learning process. However, this visual testing methodology is undesirable for various reasons, beginning with the assumption that the created drawings are proof of the person’s knowledge of the study object. Two other points of concern are the objectivity of the assessment, as it is done manually and with multiple assessors, and the long period of time between instruction and assessment, which subsequently leads to feedback with little effect.

The research was started to address these problems by developing a visual CBT test that is in alignment with the learning process, is objective, and provides feedback in a relatively short period of time. Upon closer examination of these problems, it became apparent that solely developing a test was insufficient due to a lack of knowledge of the precise education objectives – the to be learned construct – relative to visual learning and visual testing. Therefore, the research

was broadened in scope by taking a design oriented approach in studying the educational situation as a whole, based on the interaction between instruction and assessment, and with the research question: *What is an effective and efficient combination of instruction and assessment for achieving education objectives that are linked to learning through observation of concrete biological material?*

Preparatory research, including research of the literature, resulted in a conceptual framework. Subsequently, two independent lines of research were carried out and jointly contributed to answering the research question. The scope of the primary line of research was the development of a comprehensive instruction-assessment combination. The second line of research focused on finding most appropriate form of representation for the assessment of the learned construct based on two factors: degree of realism in the representation and presence/absence of contextual information.

First, a suitable instruction-assessment combination for this educational situation was sought. Hereby, the existing instruction was compared to an adjusted instruction in which drawings were no longer made. This entailed the decoupling of instruction and assessment, as assessment of drawings made during instruction was no longer a feasible testing tool. Instead, a paper-based test was developed that was electronically processed, thereby making the assessment unambiguous. The instruction was also improved through the modification of the practicum tasks; the original task of producing drawings was replaced with the task of filling in anatomical names in presented drawings. During the practicum, the students received feedback during a group discussion on the filled in drawings. Experimental comparison of the two forms of instruction showed both to be equally effective. For further research the adjusted form of instruction was selected; the adjustments made the instruction process more efficient due to time-savings in the instruction – because drawings are not made – and decrease in assessment time – because the drawings are not assessed –, and quicker feedback is given to the students as the test is marked and returned to the students in a short amount of time.

The comprehensiveness of the paper-based test was subsequently evaluated and proved to be insufficient. This lead to an evaluation and redefinition of the construct itself into three sub-constructs. In the operationalization of a CBT test, this lead to three assessment tasks: recognizing anatomical objects based on the

visual characteristics, naming these objects in Latin, and establishing relationships between the structures based on their functionality and causal behavior. The result was an automated test with 55 test items. To research the most suitable form of representation for testing the construct, the factors realism and context were operationalized resulting in four versions of the test: (1) test Photo with context, (2) test Photo without context, (3) test Model with context, and (4) test Model without context.

In an experiment, this first CBT test was evaluated to determine whether the developed instruction-assessment combination was comprehensive with respect to the construct. Additionally, it was explored whether in this format statements could be made regarding the most suitable forms of representation with the four test versions. From the results of the experiment it appeared the test could be improved both generally and at the task-level. The reliability of some test components in the four test versions, especially recognition, turned out to be insufficient and various items proved to be weak and indistinguishable. In total, 33% of the items had to be modified. Also, a first analysis of variance on the test versions to detect possible (main-) effects for the factors realism and context did not yield any effects.

The evaluation of the CBT test resulted in a revised test with which the experiment was conducted anew with four versions of the test. The results of this experiment showed a clear improvement of the assessment quality in comparison with the prior test. The experiment also demonstrated that the weakest part was the assessment task of recognition within the test versions with models rather than photos. Comparison of the four test versions indicated that they measured the same constructs globally, even though a number of items functioned differently across the various forms of assessment, and that the four test versions were equitable. With respect to the forms of representation, main-effects were found for the factor realism for the overall test and for the assessment task of recognition of objects and the task of establishing relationships between objects. In all cases, but for a limited group of items, the found main-effects favored models over photos.

The research in its entirety has led to the production of an instruction-assessment combination that is suited for education and that addresses the problems that gave rise to the research. A comprehensive CBT test was developed for the

learned construct; this test is objective due to its unambiguous automated marking and it provides immediate feedback. The added value is improvement of the quality of the instruction; the modified instruction is less time-consuming and specific attention is given to feedback during instruction. Additionally, instruction and assessment match well with a clear division between instruction and assessment, which was not the case in the drawing instruction. As far as representation format is concerned, more research is required in order to provide a conclusive answer.

The research in its current set-up did not have many candidates who participated. Administering the test multiple times will result in greater volume of data that after analysis will make a clear match between form of representation and assessment task.

As far as the implementation of the test is concerned, items that have been defined can be improved. However, that differs per test version. It is up to the teachers to improve those items that are not distinct and to construct a suitable test. Because tests must be equitable, it is possible to create a database with the items – deemed to be suitable – with which a test can be constructed.

BIJLAGEN

BIJLAGEN HORENDE BIJ HOOFDSTUK 3

BIJLAGE 3.1: PAPIEREN TOETS, TEST A & TEST B

TEST A

Voor u ligt een vragenlijst over het practicum “**Anatomie rat**”. De vraagnummers zijn in **vet** gedrukt. Bij deze vragenlijst ligt een antwoordformulier met antwoordmogelijkheden van 0 tot en met 9. Antwoord 0 niet gebruiken! Deze antwoordmogelijkheden komen overeen met de keuzemogelijkheden in de vragenlijst. De vragen hebben uitsluitend betrekking op de rat.

Vraag 1

In de halsstreek liggen 3 typen klieren, namelijk lymfeklieren, speekselklieren en traanklieren.

(LA= glandula lacrimalis, PA= glandula parotis, PR= glandula praeputialis, SL= glandula sublingualis, SM= glandula submandibularis).

In welke van de onderstaande combinaties worden 3 speekselklieren genoemd?

1. LA – SL – PR
2. LA – PA – SM
3. PA – LA – PR
4. PR – PA – LA
5. PR – LA – SM
6. SM – SL – PA
7. SM – PR – SL
8. SL – SM – LA

Vraag 2

De verteringssappen uit de pancreas bereiken de dunne darm via de _____

1. Ductus parotideus
2. Ductus choledochus
3. Ductus pancreatus
4. Ductus deferens

Vraag 3

Hoeveel longlobben zijn er links (l) en rechts (r) in de borstholte?

1. Longen (l:2, r:2)
2. Longen (l:3, r:3)
3. Longen (l:3, r:2)
4. Longen (l:2, r:3)

Vraag 4

Onderstaande afbeeldingen geven schematisch het coecum weer. Deze afbeeldingen corresponderen niet met de natuurlijke ligging in de rat. Welk schema is een correcte weergave van de toe- en afvoer van het **coecum**?

Vraag 5 - 6

Een voedselbrok komt vanuit de oesophagus de maag binnen in de (5)_____ en verlaat de maag via de (6)_____.

Maak gebruik van de onderstaande namenlijst:

1. Pylorus
2. Fundus
3. Cardia

Vraag 7

De bloedstroom in de grote en kleine bloedsomloop wordt verzorgd door het hart.

Het hart bestaat uit 4 delen, nl. 2 linker en 2 rechter delen.

Welke 2 hartdelen maken deel uit van de grote bloedsomloop?

1. Linker atrium en linker ventrikel
2. Rechter atrium en rechter ventrikel
3. Linker atrium en rechter ventrikel
4. Rechter atrium en linker ventrikel

Vraag 8

Stelling: De rat bezit een galblaas.

Deze stelling is:

1. juist
2. onjuist

Vraag 9

De mediane doorsnee door het achterlijf van de rat is schematisch weergegeven. Er zijn daarbij 3 uitmondningen getekend.

Benoem de uitmondning waar de pijl naar wijst.

Maak gebruik van de onderstaande namenlijst:

1. penis	4. uitmondning urethra
2. urogenitale opening	5. genitale opening / vagina
3. anus	

Vraag 10

Onderstaande afbeeldingen geven schematisch de nier weer. Deze afbeeldingen corresponderen niet met de natuurlijke ligging in de rat. Welk schema is een correcte weergave van de toe- en afvoer van de bloedvaten van de **nier**?

Vraag 11

Welke tekening van de buikholte is juist?

Vraag 12

Aan de ventrale zijde worden 2 incisies gemaakt, namelijk volgens de lijn A_A en de lijn B_B.

Welk onderstaand plaatje geeft het juiste transversale beeld van de oppervlakkige spieren weer bij incisie A_A?

1.

← ventrale huid

2.

← ventrale huid

3.

← ventrale huid

4.

← ventrale huid

5.

← ventrale huid

Rectum – nier

Vraag 13

Onder de huid aan de ventrale zijde komt u naar dorsaal toe, 3 ventrale oppervlakkige spieren tegen.

Benoem de spieren op de aangegeven plek in de juiste volgorde van **ventraal naar dorsaal**.

Maak daarbij gebruik van de onderstaande namenlijst:

1. Musculus pectoralis - Musculus rectus abdominus - Musculus intercostalis externus
2. Musculus deltoïdeus - Musculus rectus abdominus - Musculus intercostalis externus
3. Musculus obliquus abdominus externus - Musculus sternomastoïdeus - Musculus masseter
4. Musculus pectoralis - Musculus obliquus abdominus externus - Musculus sternomastoïdeus
5. Musculus deltoïdeus - Musculus sternomastoïdeus - Musculus obliquus abdominus internus
6. Musculus obliquus abdominus externus - Musculus rectus abdominus - Musculus obliquus abdominus internus

Vraag 14

De kleine bloedsomloop bestaat uit 2 hartdelen, de longen en enkele aan- en afvoerende bloedvaten.

Welke bloedvaten maken deel uit van de kleine bloedsomloop?

1. Aorta en vena azygos
2. Arteria pulmonalis en vena azygos
3. Aorta en vena cava rostralis
4. Arteria pulmonalis en vena cava rostralis
5. Aorta en vena pulmonalis
6. Arteria pulmonalis en vena pulmonalis

Vraag 15

Behoort de vagina tot het oviduct?

1. Ja
2. Nee

Vraag 16

De verbindingen van de aorta (A) met de nieren (N) en de bijnieren (B) zijn getekend, evenals de verbindingen van de ovaria (O) met de ongepaarde vagina (V). Welke schematische afbeelding is een correcte ventrale weergave van de organen van de vrouwtjes rat?

1.

Rostraal

2.

Caudaal

3.

Rostraal

4.

Caudaal

Vraag 17

Welke tekening van de rechterhelft van de halsstreek geeft het juiste beeld weer van de speekselklieren, lymfeklieren en vena jugularis externa?

Vraag 18

De voedselbrok verlaat de maag en doorloopt dan alle delen van het darmstelsel.

Wat is in normale gevallen de juiste volgorde?

((C= Colon, D= Duodenum, I= Ileum, R= Rectum)

1. C - D - R - I
2. D - I - C - R
3. I - C - R - D
4. R - C - D - I
5. C - D - I - R
6. D - R - I - C
7. I - R - D - C
8. R - I - C - D

Vraag 19

Hoeveel tenen heeft de rat aan de voorpoot en aan de achterpoot?

1. Voorpoot 4; achterpoot 4
2. Voorpoot 5; achterpoot 5
3. Voorpoot 4; achterpoot 5
4. Voorpoot 5; achterpoot 4

Vraag 20

Zit de **plexus pampiniformis** in de bloedvoorziening naar de caudale zijde van de testis of naar de rostrale zijde?

1. Caudaal
2. Rostraal

Vraag 21

Onderstaand schema is een overzicht van het **mannelijk geslachtsorgaan**.

Waar mondt de glandula prostata uit?

1. op locatie 0	4. Op locatie 3
2. op locatie 1	5. Op locatie 4
3. op locatie 2	

TEST B

Voor u ligt een vragenlijst over het practicum "**Anatomie rat**". De vraagnummers zijn in **vet** gedrukt. Bij deze vragenlijst ligt een antwoordformulier met antwoordmogelijkheden van 0 tot en met 9. Deze antwoordmogelijkheden komen overeen met de keuzemogelijkheden in de vragenlijst. Antwoord 0 niet gebruiken. De vragen hebben uitsluitend betrekking op de rat.

Vraag 22

Afbeelding 4 geeft de **situs viscerum borstholte** weer. Welk genummerd onderdeel is een weergave van de **thymus**?

Vraag 23

In de **plexus pampiniformis** treden een arteria en vena in contact met elkaar door een sterke kronkeling van de vena om de arteria. Wordt de plexus pampiniformis gevormd door de arteria en vena spermatica externa of door de arteria en vena spermatica interna?

1. Arteria en vena spermatica interna
2. Arteria en vena spermatica externa

Vraag 24

Afbeelding 5 geeft de **grote bloedvaten en organen van de hals** weer. Welk genummerd onderdeel is een weergave van de **nervus vagus**?

Vraag 25

Afbeelding 2 geeft de **buikholte** van de **vrouwtjes rat** weer. Benoem onderdeel 1.

Maak een keuze uit de volgende namenlijst:

1. Ureter	4. Uterus
2. Tuba	5. Blaas
3. Ductus urogenitalis	

Vraag 26

De mediane doorsnee door het achterlijf van de rat is schematisch weergegeven. Er zijn daarbij 3 uitmondingen getekend.

Benoem de plek waar de pijl naar wijst.

Maak gebruik van de onderstaande namenlijst:

1. penis	4. uitmonding urethra
2. urogenitale opening	5. genitale opening / vagina
3. anus	

Vraag 27

Hieronder staan 2 tekeningen van het ventrale aanzicht van het hart weergegeven.

Welke van de 2 tekeningen is een correcte weergave?

Vraag 28

Mondt de vena azygos uit in de vena cava caudalis of in de vena cava rostralis sinistra ?

1. Vena cava caudalis
2. Vena cava rostralis sinistra

Vraag 29

Afbeelding 1 geeft de **situs viscerum buikholte** weer. Welk genummerd onderdeel is een weergave van het **coecum**?

Vraag 30

Afbeelding 3 geeft de **kop en hals** ventraal weer.

Benoem onderdeel 2.

Maak een keuze uit de volgende namenlijst:

1. Vena cava rostralis	5. Vena cava caudalis
2. Larynx	6. Trachea
3. Oesophagus	7. Arteria carotis
4. Nervus vagus	8. Vena jugularis externa

Vraag 31

Afbeelding 2 geeft de **buikholte van de vrouwtjes rat** weer. Benoem onderdeel 2.

Maak een keuze uit de volgende namenlijst:

1. Ureter	4. Uterus
2. Tuba	5. Blaas
3. Ductus urogenitalis	

Vraag 32

Afbeelding 4 geeft de **situs viscerum borstholte** weer. Welk genummerd onderdeel is een weergave van het **diaphragma**?

Vraag 33

Onderstaande afbeeldingen geven schematisch de maag weer. Deze afbeeldingen corresponderen niet met de natuurlijke ligging. Welk schema is een correcte weergave van de toe- en afvoer van de **maag**?

Vraag 34

Op de overgang van welke 2 delen van het darmstelsel ligt het **coecum**?

((C= Colon, D= Duodenum, I= Ileum en R= Rectum)

1. C - R
2. D - C
3. I - C
4. R - D
5. C - D
6. D - I
7. I - R
8. R - I

Vraag 35

Onderstaand schema is een overzicht van het **mannelijk geslachtsorgaan**.

Waar ligt de caput epididymidis?

1. op locatie 0	4. Op locatie 3
2. op locatie 1	5. Op locatie 4
3. op locatie 2	

Vraag 36

Aan de ventrale zijde worden 2 incisies gemaakt, namelijk volgens de lijn A_A en de lijn B_B.

Welk onderstaand plaatje geeft het juiste transversale beeld van de oppervlakkige spieren weer bij incisie B_B?

1. ← ventrale huid
2. ← ventrale huid
3. ← ventrale huid
4. ← ventrale huid
5. ← ventrale huid

Raster = spier

Vraag 37

Afbeelding 3 geeft de **kop en hals** ventraal weer. Benoem onderdeel 1.

Maak gebruik van de onderstaande namenlijst:

1. Musculus sternomastoïdeus	6. Musculus deltoïdeus
2. Musculus pectoralis	7. Musculus masseter
3. Musculus intercostalis externus	8. Musculus rectus abdominus
4. Musculus digastricus	9. Musculus obliquus abdominis internus
5. Musculus obliquus abdominis externus	

Vraag 38

Afbeelding 1 geeft de **situs viscerum buikholte** weer. Welk genummerd anatomisch onderdeel is een weergave van de **milt**?

Vraag 39

Afbeelding 5 geeft de **grote bloedvaten en organen van de hals** weer. Welk genummerd onderdeel is een weergave van de **arteria carotis**?

Vraag 40

Welke van de 2 tekeningen van de ureters, de vesica urinaria en de urethra is correct?

1.

2.

Vraag 41

Onder de huid aan de ventrale zijde komt u naar dorsaal toe, 3 ventrale oppervlakkige spieren tegen.

Benoem de spieren op de aangegeven plek in de juiste volgorde van **ventraal naar dorsaal**.

Maak daarbij gebruik van de onderstaande namenlijst:

1. Musculus pectoralis - Musculus rectus abdominus - Musculus intercostalis externus
2. Musculus deltoïdeus - Musculus rectus abdominus - Musculus intercostalis externus
3. Musculus obliquus abdominus externus - Musculus sternomastoïdeus - Musculus masseter
4. Musculus pectoralis - Musculus obliquus abdominus externus - Musculus sternomastoïdeus
5. Musculus deltoïdeus - Musculus sternomastoïdeus - Musculus obliquus abdominus internus
6. Musculus obliquus abdominus externus - Musculus rectus abdominus - Musculus obliquus abdominus internus

BIJLAGE 3.2 REPRESENTATIEVROEMEN HORENDE BIJ DE PAPIERENTOETS, VERSIE 1 EN VERSIE 2

Representatievorm versie 1

BIJLAGEN HORENDE BIJ HOOFDSTUK 4

BIJLAGE 4.1: CBT-TOETS 1STE VERSIE

Vraag	Type itemformat	Vraagstelling	Visualisatievorm voorbeelden
			<i>Toetsvorm: Foto zonder context</i>
		Thema 1: Buikholte en spijsvertering	
1	herkennen	Welk plaatje is de juiste weergave van het ventraal aanzicht van de rattemaag?	
2	benoemen	Benoem het aangegeven onderdeel van de rattemaag.	
3	herkennen	Welke klier is een weergave van de glandula submandibularis ?	
4	relatie	Geef door het verbinden van de verschillende organen de richting van het voedseltransport weer vanaf de maag, uitgaande van een duodenum van 12 cm .	
5	relatie	Geef door het verbinden van de verschillende organen de richting van het voedseltransport weer vanaf het coecum, uitgaande van een duodenum van 12 cm .	
6	benoemen	Benoem het aangegeven onderdeel van de rattemaag.	
7	herkennen	Welke klier is een weergave van de glandula lacrimalis ?	

Vraag	Type itemformat	Vraagstelling	Visualisatievorm voorbeelden
8	herkennen	Welk plaatje is de juiste weergave van het duodenum?	
9	herkennen	Welk plaatje is een juiste weergave van het spijsverteringskanaal van maag tot en met rectum (het groen omkaderd deel is het duodenum met een lengte van 12 cm)?	
10	relatie	In het onderstaande schema zijn enkele anatomische structuren weergegeven of benoemd die bijdragen aan de afbraak van voedsel en aan de opname van stoffen in het bloed. Verbind deze structuren met elkaar zoals jij denkt dat ze in functioneel verband met elkaar staan.	
11	benoemen	Benoem het aangegeven onderdeel van de rattemaag.	
12	herkennen	Welk deel van de kop is een weergave van de musculus masseter ?	
		Thema 2: Spielen	
13	herkennen	Welk deel van de spieren is een weergave van de musculus deltoïdeus ?	
14	herkennen	Welk deel van de spieren is een weergave van de musculus rectus abdominus ?	
			Toetsvorm: Schema zonder context

Vraag	Type itemformat	Vraagstelling	Visualisatievorm voorbeelden
15	herkennen	Welk deel van de spieren is een weergave van de musculus rectus abdominus ?	
16	herkennen	Welk deel van de spieren is een weergave van de musculus obliquus abdominus externus ?	
17	herkennen	Welk deel van de spieren is een weergave van de musculi intercostales externi ?	
18	herkennen	Welk deel van de spieren is een weergave van de musculus obliquus abdominus internus ?	
		Thema 3: Geslachtsorganen	
19	benoemen	Benoem het aangegeven anatomisch onderdeel.	
20	relatie	Geef in de onderstaande afbeeldingen het transport van de onbevvruchte eicel weer.	
21	benoemen	De rechtertestis is afgebeeld. Benoem het aangegeven anatomisch onderdeel.	
22	benoemen	De rechtertestis is afgebeeld. Benoem het aangegeven anatomisch onderdeel.	IDEM

Vraag	Type itemformat	Vraagstelling	Visualisatievorm voorbeelden
23	herkennen	Welk plaatje is een juiste weergave van de rechtertestis en diens afvoergang?	
24	benoemen	Benoem het aangegeven anatomisch onderdeel.	
25	benoemen	Benoem het aangegeven anatomisch onderdeel.	IDEM
26	relatie	Kies uit de onderdelen van het mannelijk urogenitaalstelsel (rechterdeel) de correcte afbeeldingen en geef hun functionele relaties aan.	
27	benoemen	Benoem de aangegeven klier van de rechter mannelijke geslachtsorganen.	
28	benoemen	Benoem de aangegeven klier van de rechter mannelijke geslachtsorganen.	IDEM
			<i>Toetsvormorm: Foto en context</i>
29	herkennen	Welk plaatje is een juiste weergave van het mannelijk urogenitaalstelsel (rechterdeel)?	
30	benoemen	Benoem het aangegeven anatomisch onderdeel.	
31	benoemen	Benoem het aangegeven anatomisch onderdeel.	
32	relatie	Geef in de onderstaande afbeelding het transport van de spermatozoïden weer.	

Vraag	Type itemformat	Vraagstelling	Visualisatievorm voorbeelden
33	relatie	Geef in de onderstaande afbeeldingen het transport van de spermatozoïden weer.	
		Thema 4: Borstholte en grote bloedvaten	
34	herkennen	Welk plaatje is een correcte weergave van de vertakkingen van de venae pulmonales naar de longlobben vanuit ventraal aanzicht?	
35	herkennen	Welk plaatje is een weergave van de grote lichaamsvenen in relatie tot het hart? (ventraal aanzicht)	
36	benoemen	Benoem het aangegeven anatomisch onderdeel.	
37	benoemen	Benoem het aangegeven anatomisch onderdeel.	IDEM
38	benoemen	Benoem het aangegeven anatomisch onderdeel.	IDEM
39	herkennen	Welk plaatje is een juiste weergave van de vertakkingen van de aortaboog? (ventraal aanzicht)	
40	benoemen	Benoem het aangegeven anatomisch onderdeel.	
41	benoemen	Benoem het aangegeven anatomisch onderdeel.	IDEM
42	benoemen	Benoem het aangegeven anatomisch onderdeel.	IDEM

Vraag	Type itemformat	Vraagstelling	Visualisatievorm voorbeelden
			<i>Toetsvorm:: Schema en context</i>
43	benoemen	Benoem het aangegeven anatomisch onderdeel.	
44	benoemen	Benoem het aangegeven anatomisch onderdeel.	IDEM
45	relatie	Geef aan op welke delen van het hart de verschillende onderdelen zijn aangesloten.	
46	herkennen	Welk plaatje is een correcte weergave van het hart met aan- en afvoerende bloedvaten? (ventraal aanzicht)	
47	relatie	Geef van de bloedvatuiteinden die rechtstreeks verbonden zijn met de rechterventrikel van het hart dmv een pijl de bloedstroomrichting aan. (ventraal aanzicht)	
48	relatie	Geef van de bloedvatuiteinden die rechtstreeks verbonden zijn met de linkeratrium van het hart dmv een pijl de bloedstroomrichting aan. (ventraal aanzicht)	IDEM
49	relatie	Geef van de bloedvatuiteinden die rechtstreeks verbonden zijn met de rechteratrium van het hart dmv een pijl de bloedstroomrichting aan. (ventraal aanzicht)	IDEM
50	relatie	Geef van de bloedvatuiteinden die rechtstreeks verbonden zijn met de linkerventrikel van het hart dmv een pijl de bloedstroomrichting aan. (ventraal aanzicht)	IDEM

Vraag	Type itemformat	Vraagstelling	Visualisatievorm voorbeelden
		Thema (Bij)nieren en bloedvaten	
51	relatie	Rostraal zijn de bloedstroomrichtingen in de lichaamsslagader en de onderste holle ader gegeven. (ventraal aanzicht) Verbindt op de juiste wijze – denk aan de bloedstroomrichting – deze bloedvaten met de rechter nier en met de rechter achterpoot.	
52	relatie	Rostraal zijn de bloedstroomrichtingen in de lichaamsslagader en de onderste holle ader gegeven. (ventraal aanzicht) Geef de bloedstroomrichting aan van de arteria iliaca en de vena iliaca.	IDEM
53	benoemen	In het onderstaande plaatje wordt de stroomrichting van het bloed in de arteriën en venen aangegeven (rode pijlen). Benoem de met een ster aangegeven arteriën en venen.	
54	herkennen	Welk onderdeel van de nier is een weergave van de medulla ?	
55	herkennen	Welk onderdeel van de nier is een weergave van de papillus ?	IDEM

BIJLAGEN HORENDE BIJ HOOFDSTUK 5

BIJLAGE 5.1: CBT-TOETS 2DE VERSIE

Vraag	Type itemformat	Vraagstelling	Visualisatievorm voorbeelden
			Toetsvorm: Foto zonder context
		Thema 1: Buikholte en spijsvertering	
Q01	Herkennen	Welk plaatje is de juiste weergave van het ventraal aanzicht van de rattemaag?	
Q02	Benoemen	Benoem het aangegeven onderdeel van de rattemaag. (ventraal aanzicht)	
Q03	Herkennen	In de afbeelding worden de verschillende klieren van de kop met een pijl aangegeven. Welke klier is een weergave van de glandula submandibularis ? (ventraal aanzicht)	
Q04	Relatie	Ten behoeve van de spijsvertering wordt voedsel getransporteerd door verschillende organen. Verbind de maag met het daarop volgend orgaan, uitgaande van een duodenum van 12 cm . (ventraal aanzicht)	
Q05	Relatie	Ten behoeve van de spijsvertering wordt voedsel getransporteerd door verschillende organen. Verbind het duodenum met het daarop volgend orgaan, uitgaande van een duodenum van 12 cm . (ventraal aanzicht)	
Q06	Relatie	Ten behoeve van de spijsvertering wordt voedsel getransporteerd door verschillende organen. Verbind het ileum met het daarop volgend orgaan, uitgaande van een duodenum van 12 cm . (ventraal aanzicht)	
Q07	Relatie	Ten behoeve van de spijsvertering wordt voedsel getransporteerd door verschillende organen. Verbind het coecum met het daarop volgend orgaan, uitgaande van een duodenum van 12 cm . (ventraal aanzicht)	
Q08	Benoemen	Benoem het aangegeven onderdeel van de rattemaag. (ventraal aanzicht)	
Q09	Herkennen	In de afbeelding worden de verschillende klieren van de kop met een pijl aangegeven. Welke klier is een weergave van de glandula lacrimalis ? (ventraal aanzicht)	

Vraag	Type itemformat	Vraagstelling	Visualisatievorm voorbeelden
Q10	Herkennen	Welk plaatje is een juiste weergave van het spijsverteringskanaal van maag tot en met rectum (het groen omkaderd deel is het duodenum met een lengte van 12 cm)? (ventraal aanzicht)?	
Q11	Relatie	Voedsel wordt verteerd in het spijsverteringskanaal. Voor dit proces worden er stoffen aan het spijsverteringskanaal afgegeven . Geef de juiste route weer van deze stoffen door de anatomische structuren met elkaar te verbinden. (ventraal aanzicht)	
Q12	Relatie	Voedsel wordt verteerd in het spijsverteringskanaal. T.g.v. dit proces worden er stoffen in het bloed opgenomen die in het lichaam verbruikt worden. Geef de juiste route weer van deze stoffen door de anatomische structuren met elkaar te verbinden. (ventraal aanzicht)	
Q13	Benoemen	Benoem het aangegeven onderdeel van de rattemaag. (ventraal aanzicht)	
Q14	Herkennen	In de afbeelding worden de verschillende spieren van de kop met een pijl aangegeven. Welk deel van de kop is een weergave van de musculus masseter ? (ventraal aanzicht)	
			Toetsvorm: Schema zonder context
		Thema 2: Spieren	
Q15	Herkennen	In de afbeelding worden de verschillende spieren met een pijl aangegeven. Welk deel van de spieren is een weergave van de musculus deltoïdeus ? (ventraal aanzicht)	
Q16	Herkennen	In de afbeelding worden de verschillende spieren met een pijl aangegeven. Welk deel van de spieren is een weergave van de musculus rectus abdominus ? (ventraal aanzicht)	
Q17	Herkennen	In de afbeelding worden de verschillende spieren met een pijl aangegeven. Welk deel van de spieren is een weergave van de musculus rectus abdominus ? (ventraal aanzicht)	
Q18	Herkennen	In de afbeelding worden de verschillende spieren met een pijl aangegeven. Welk deel van de spieren is een weergave van de musculus obliquus abdominus externus ? (ventraal aanzicht)	

Vraag	Type itemformat	Vraagstelling	Visualisatievorm voorbeelden
Q19	Herkennen	In de afbeelding worden de verschillende spieren met een pijl aangegeven. Welk deel van de spieren is een weergave van de musculi intercostales externi ? (ventraal aanzicht)	
Q20	Herkennen	In de afbeelding worden de verschillende spieren met een pijl aangegeven. Welk deel van de spieren is een weergave van de musculus obliquus abdominus internus ? (ventraal aanzicht)	
Thema 3: Geslachtsorganen			
Q21	Benoemen	Benoem het aangegeven anatomisch onderdeel. (ventraal aanzicht)	
Q22	Relatie	Geef in de onderstaande afbeelding het transport van de onbevuchte eicel weer tot de plaats waar de bevruchting plaatsvindt. (ventraal aanzicht)	
Q23	Benoemen	De rechtertestis is afgebeeld. Benoem het aangegeven anatomisch onderdeel. (ventraal aanzicht)	
Q24	Benoemen	De rechtertestis is afgebeeld. Benoem het aangegeven anatomisch onderdeel. (ventraal aanzicht)	
Q25	Herkennen	Welk plaatje is een juiste weergave van de rechtertestis en diens afvoergang? E= deel epididymis, *= ductuli efferentes (ventraal aanzicht)	
Q26	Benoemen	Benoem het aangegeven anatomisch onderdeel. E= deel epididymis, *= ductuli efferentes (ventraal aanzicht)	
Q27	Benoemen	Benoem het aangegeven anatomisch onderdeel. E= deel epididymis, *= ductuli efferentes (ventraal aanzicht)	
Q29	Benoemen	Benoem de aangegeven klier van de rechter mannelijke geslachtsorganen. (ventraal aanzicht)	

Vraag	Type itemformat	Vraagstelling	Visualisatievorm voorbeelden
			Toetsvorm: Foto met context
Q30	Benoemen	Benoem de aangegeven klier van de rechter mannelijke geslachtsorganen. (ventraal aanzicht)	
Q31	Herkennen	Welk plaatje is een juiste weergave van het mannelijk urogenitaalstelsel (rechterdeel)? (ventraal aanzicht) P= penis	
Q32	Benoemen	Benoem het aangegeven anatomisch onderdeel? (ventraal aanzicht)	
Q33	Benoemen	Benoem het aangegeven anatomisch onderdeel? (ventraal aanzicht)	
Q34	Relatie	Geef in de onderstaande afbeelding het transport van de spermatozoïden weer. (ventraal aanzicht)	
Q35	Relatie	Geef in de onderstaande afbeeldingen het transport van de spermatozoïden weer naar de plek van de bevruchting. (ventraal aanzicht)	
		Thema 4: Borstholte en grote bloedvaten	
Q36	Herkennen	Welk plaatje is een correcte weergave van de vertakkingen van de venae pulmonales naar de longlobben vanuit ventraal aanzicht? L= links, R= rechts, *= aansluiting bloedvat op het hart	
Q37	Herkennen	Welk plaatje is een correcte weergave van de grote lichaamsvenen in relatie tot het hart? (ventraal aanzicht) groene lijn= mediaan; *= aansluiting bloedvat op het hart, VA= vena azygos	
Q38	Benoemen	Benoem het aangegeven anatomisch onderdeel. (ventraal aanzicht)	
Q39	Benoemen	Benoem het aangegeven anatomisch onderdeel. (ventraal aanzicht)	

Q40	Benoemen	Benoem het aangegeven anatomisch onderdeel. (ventraal aanzicht)	
Q41	Herkennen	Welk plaatje is een juiste weergave van de vertakkingen van de aortaboog? (ventraal aanzicht) groene lijn= mediaan; *= aansluiting bloedvat op het hart	
Q42	Benoemen	Benoem het aangegeven anatomisch onderdeel. (ventraal aanzicht)	
Q43	Benoemen	Benoem het aangegeven anatomisch onderdeel. (ventraal aanzicht)	
Q44	Benoemen	Benoem het aangegeven anatomisch onderdeel. (ventraal aanzicht)	
			<i>Toetsvorm: Schema met context</i>
Q45	Benoemen	Benoem het aangegeven anatomisch onderdeel. (ventraal aanzicht)	
Q46	Relatie	Geef aan op welke deel van het hart de grote lichaamsvenen zijn aangesloten. VA= vena azygos, VCC= vena cava caudalis, *= aansluiting bloedvat op het hart. (ventraal aanzicht)	
Q47	Relatie	Geef aan op welke deel van het hart de venae pulmonales zijn aangesloten. VP= vena pulmonalis, *= aansluiting bloedvat op het hart. (ventraal aanzicht)	
Q48	Relatie	Geef aan op welke deel van het hart de aortaboog is aangesloten. AC= arteria carotis, AD= aorta dorsalis, TB= truncus brachiocephalicus, *= aansluiting bloedvat op het hart. (ventraal aanzicht)	
Q49	Herkennen	Welk plaatje is een correcte weergave van het hart met aan- en afvoerende bloedvaten? (ventraal aanzicht)	

Q50	Relatie	Geef van de bloedvatuiteinden die rechtstreeks verbonden zijn met de rechterventrikel van het hart dmv een pijl de bloedstroomrichting aan. (ventraal aanzicht)	
Q51	Relatie	Geef van de bloedvatuiteinden die rechtstreeks verbonden zijn met de linkeratrium van het hart dmv een pijl de bloedstroomrichting aan. (ventraal aanzicht)	
Q52	Relatie	Geef van de bloedvatuiteinden die rechtstreeks verbonden zijn met de linkerventrikel van het hart dmv een pijl de bloedstroomrichting aan. (ventraal aanzicht)	
Q53	Functie	Geef van de bloedvatuiteinden die rechtstreeks verbonden zijn met de rechteratrium van het hart dmv een pijl de bloedstroomrichting aan. (ventraal aanzicht)	
		Thema 5: (Bij)nieren en bloedvaten	
Q54	Relatie	Vanuit ventraal aanzicht zijn de lichaamsslagader en de onderste holle ader weergegeven. Verbind op de juiste wijze – denk aan de bloedstroomrichting – deze bloedvaten met de rechter nier en met de rechter achterpoot.	
Q55	Relatie	Geef de bloedstroomrichting aan van de arteria iliaca en de vena iliaca. (ventraal aanzicht)	
Q56	Benoemen	In het onderstaand plaatje wordt de stroomrichting van het bloed in de arteriën en venen aangegeven (rode pijlen). Benoem de met een ster aangegeven arteriën en venen. (ventraal aanzicht)	
Q57	Herkennen	Welk onderdeel van de nier is een weergave van de medulla ? (ventraal aanzicht)	
Q58	Herkennen	Welk onderdeel van de nier is een weergave van de papillus ? (ventraal aanzicht)	

BIJLAGEN HORENDE BIJ HOOFDSTUK 6

Bijlage 6.1 *Protocolanalyse met trefwoorden*

Gezocht aspect	Gebruikte trefwoorden
Duidelijk	Duidelijk, maar niet “niet duidelijk” of “onduidelijk”
Onduidelijk/ niet duidelijk	onduidelijk; niet duidelijk
Vaag plaatje	vaag plaatje; vage plaatje
Onduidelijke of moeilijke vraag	Geen trefwoorden maar met de hand gescoord
Kleuren	rode;rood;bruin;groen;kleur;gelig
Locatie/oriëntatie	link;recht;komt uit;vanuit;uitkom;split;in de buurt;hoog ;laag ;orient;oriënt;boven (mét spatie, waardoor “bovendien” niet wordt geteld);midden; onder (maar niet “onder andere”)
Vorm	vorm;groot;grote;klein
Syllabus/handleiding	syllabus;handleiding
Voorsnijden/voorbepreking	voorbeprek;voorsnijde
Dissectie/ gezien in practicum /tekening	dissectie;gezien;tekening;practicu;snij (maar niet “voorsnijd”)
Nabespreking	nabesprek;poster
Begeleiding	assistent;angelic;uitleg
Gok	gok
Beschrijving met elementen uit de leerstof	Deod, coec, colo, ventrik, boez, aorta, venen, vena, bloed, zuurstof, maag, klier, darm, ader, hart, ileum, spier, zaad, eikel, long, bloedvat

Bijlage 6.2 Totaal overzicht van de classificatie van gescoorde responses

	Totalen voor			
	<i>herkennen</i>	<i>benoemen</i>	<i>relaties</i>	<i>alles</i>
	841	887	773	2501
Totaal duidelijk	18	13	10	41
onduidelijk / niet duidelijk	7	4	11	22
vaag plaatje	0	1	1	2
vage plaatjes	1	0	1	2
Totaal onduidelijk / vaag	8	5	13	26
Rij met verwijzing naar duidelijkheid	26	17	23	66
Onduidelijke plaatjes / foto's	5	1	5	11
Onduidelijke of moeilijke vraag (met de hand gescoord)	1	1	9	11
Totaal kleuren	14	17	4	35
Totaal locatie/ oriëntatie	214	128	132	474
Totaal vorm	72	62	33	167
Rij met verwijzing naar vormkenmerken	281	182	162	625
Totaal syllabus / handleiding	23	37	19	79
Totaal voorsnijden / voorbespreking	15	15	16	46
Totaal dissectie / gezien in practicum /tekening	242	199	168	609
Totaal nabespreking	51	44	26	121
Totaal begeleiding	106	73	73	252
Rij met verwijzing naar instructiecomponent	351	313	257	921
Beschrijvend met elementen van de leerstof	250	205	325	780
Gok	62	106	71	239
Niet-gescoorde niet-lege rijen	157	240	173	570
Wel gescoord	684	647	600	1931
Wel gescoord in procenten	81.3	72.9	77.6	77.2